

American Housing Survey

STREAMLINING THE AMERICAN HOUSING SURVEY



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



Visit PD&R's Web Site

www.huduser.org

to find this report and others sponsored by

HUD's Office of Policy Development and Research (PD&R).

Other services of HUD USER, PD&R's Research Information Service, include listservs; special interest reports, bimonthly publications (best practices, significant studies from other sources); access to public use databases; hotline 1-800-245-2691 for help accessing the information you need.

American Housing Survey

**STREAMLINING THE
AMERICAN HOUSING
SURVEY**

Prepared for:

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

Prepared by:

Frederick J. Eggers
Econometrica, Inc.
Bethesda, Maryland

June 2009

The contents of this report are the views of the contractor, and do not necessarily reflect the views or policies of the U.S. Department of Housing and Urban Development or the U.S. Government.

Table of Contents

1. Purpose.....	1
2. Approach	1
3. Recommendations	3
4. Mortgage-Related Variables	7
5. Neighborhood Conditions	9
6. Assisted Housing.....	11
7. Housing Costs, Including Utilities	12
8. Housing Features and Amenities	15
9. Housing Quality.....	16
10. Household Composition.....	18
11. Recent Movers.....	19
12. Commuting	20
13. Mobile Homes.....	21
14. Lead Based Paint	22
15. Upgrades and Remodeling.....	22

Streamlining the American Housing Survey

1. Purpose

Since 1973, the U.S. Department of Housing and Urban Development (HUD) and the Census Bureau have teamed to produce the American Housing Survey (AHS). The recently released 2007 AHS is the 22nd report on the characteristics and condition of the national housing stock; on the costs of housing, including the financing of owner-occupied housing; and on the households who occupy the stock. Data from the AHS allow researchers and policy analysts, both inside and outside the government, to document housing problems and to evaluate the operation of the housing market and of policies designed to improve housing.

The national AHS collects information from over 50,000 households every two years using an instrument that often takes over an hour to administer. The public data files contain thousands of fields. The size of the sample and the length of the questionnaire make the AHS an expensive survey. The length of the questionnaire and the AHS's longitudinal design—the same housing units are surveyed every two years—impose significant burdens on respondents. As the AHS's sponsor, HUD feels obligated to ensure that the information produced justifies the costs and burdens.

For this reason, HUD requested Econometrica, Inc., to assess the analytical usefulness of the AHS's content. Specifically, HUD asked the authors to:

Examine the AHS instrument and data set to identify elements that have some or all of these characteristics:

- *They have very little variation in value, either cross-sectionally or longitudinally.*
- *They exhibit unreasonable year-to-year changes for the same household or housing unit.*
- *They are used by few researchers, and those uses are not of notable scientific or policy importance.*

The goal is to streamline the AHS “by eliminating or modifying content that is currently of marginal usefulness.” Streamlining would reduce the cost to the government and, more importantly, would lessen the burden imposed on the public. Streamlining also could create opportunities to add new and more policy-important content to the AHS.

2. Approach

The authors undertook this task cautiously because the AHS is an immensely valuable data source. Analysts use it for a wide variety of purposes. While many of the uses are documented in the professional literature, other important uses by staff in federal agencies or in Congress or congressional agencies remain generally unpublicized. The AHS's 35-year history allows users to trace changes in housing-related variables over a period during which the U.S. population

grew from just over 200 million to over 300 million. The costs from a Type I error—keeping an unimportant or problematic variable—are small compared with the costs of a Type II error—eliminating an important variable and breaking the continuity of information it provides.

In producing this report, the authors performed the following tasks:

1. Talked to known users of the AHS to identify questionable variables and to gain their perspective on streamlining the AHS.
2. Reviewed 89 articles and papers that used the AHS and that were published between 2004 and 2007.
3. “Googled” the phrase “according to the American Housing Survey” and found 119 hits.
4. Obtained an analysis from the Census Bureau on the number of “hits” for individual AHS tables in the 6 months from October 2007 to March 2008.
5. Systematically reviewed all variables to determine the number of cases with useable values and the prevalence of edits and allocations.
6. Systematically reviewed all variables to determine the consistency of responses by the same households between 2003 and 2005.
7. Systematically reviewed all variables to determine the extent of variation between 1997 and 2005 and within 2005 across the four Census regions; across urban, suburban, and rural areas; between non-Hispanic white and all others; and between below-median income and above-median income households.

While each of the analyses produced useful insights, the combined effect is a mountain of information that is difficult to synthesize. In the end, the authors decided to focus on groups of variables as the most helpful way to interpret the information already collected. This approach also offers the best prospects for finding ways to streamline the AHS. Eliminating a variable here and there would have little impact on either the costs or burdens associated with the AHS. Eliminating groups of variables could both achieve cost and burden reductions and afford the opportunity to add new and more useful content to the AHS.

The next section contains our recommendations for ways to improve the AHS through eliminating or refining existing variables. The remaining sections discuss in greater detail the work we performed. Each section deals with a group of AHS variables; the sections correspond in general to the sections in the most recent AHS codebook. The discussion in these sections covers some of the same points included among our recommendations in Section 3, but also includes what we learned about variables that we are not recommending for elimination or refinement.

While the AHS devotes substantial time in the interviews to collecting information on income, the income-related variables are not assessed in this report. Income is one of the most

fundamental statistics collected in any survey, and therefore, the issue is not whether to collect information on income, but how to do so effectively. HUD and the Census Bureau reworked the income variables prior to the 2005 AHS national survey and have extensively analyzed the performance of the new variables.

3. Recommendations

Our recommendations fall into three classes. First, we identify a number of variables that HUD and the Census Bureau should consider eliminating from the AHS. Second, we identify two classes of variables—those associated with the financing of owner-occupied housing and with the identification of assisted housing—that definitely merit the cost and burden of collecting the information, but that also have well-recognized problems that HUD and the Census Bureau need to solve. Third, we discuss variables related to the use and costs of utilities, a group that may need further improvement and perhaps some trimming.

Variables that HUD and the Census Bureau should consider eliminating to streamline the AHS

There are some variables that we believe are so plagued with consistency problems or other issues that HUD and the Census Bureau should consider eliminating them from the AHS.

- **“X” variables:** The AHS public use file (PUF) contains a class of variables that are coded to identify only those cases that result in positive responses to a question.¹ A number of these variables showed remarkably poor consistency between responses in 2003 and 2005. We discuss three classes of these variables.
 - **Neighborhood variables:** For variables identifying problems, one would not necessarily expect consistency in answers between successive surveys.² But even in these cases, the very low levels of consistency were surprising. There are two alternatives: The inconsistency represents a problem with the variable, or the inconsistency reflects real changes between surveys. Even in the latter case, one can question the usefulness of the information provided. For example, only 10 percent of households who cited BADPER (people in neighborhood are bothersome) as a problem in 2003 did so again in 2005. If we assume that the 2003 and 2005 responses are both correct, then BADPER does not appear to provide a useful description of conditions in the neighborhood since the problem

¹ Most of these variables record an “X” in the PUF when the respondent gives a positive answer. Some record a “1” instead of an “X”, and a few record a “2”. For all of these X-variables, frequency distributions will show three values: “Blank” for missing information; “B” for not applicable; and either “X”, “1”, or “2” to indicate a positive response. If an “X” or a “1” indicates a positive response, there will be no “2” or some other value to indicate a negative response.

² Actually, for the variables that allow yes/no, this expectation is generally not correct. For example, HOLES (holes in floor) has identical answers in 2003 and 2005 for the same households in 99 percent of the cases. This high level of consistency probably relates to the general high-quality level of the stock. In 2005, only 0.8 percent of occupied units had holes in the floor (HOLES). If the same percentage applied in 2003 and the distribution of units with holes was random in both years, one would expect identical answers to HOLES in 98 percent of the cases.

(bothersome people) appears to be ephemeral. Here is a list of other neighborhood X-variables that displayed worrisome consistency:

- **BADSERV:** poor city/county services are bothersome; 11 percent of the respondents from the same households who answered “yes” in 2003 did so again in 2005.
 - **BADPRP:** undesirable neighborhood/property bothersome, 31 percent.
 - **LITTER:** litter in neighborhood bothersome, 15 percent.
 - **NOISE:** noise in neighborhood is bothersome, 12 percent.
 - **OTHNHD:** other problems are bothersome in neighborhood, 24 percent.
- **“BILL” variables:** The AHS includes 30 variables that indicate whether a given utility is billed in conjunction with another utility. For example, BILLGE tells whether gas is billed with electricity. These variables take the value of “X” (or sometimes “2”) if the answer is “yes”; “B” (not applicable) and “blank” (not reported) are the only other values allowed. BILLW (water billed with other utility) has the largest number of positive responses, 1,525. The average number of “yeses” is 173. The Census Bureau does not allocate values to any of these variables, but edits are frequent, ranging from 43,358 cases with edits to 30 cases. For every variable, the number of edits exceeds the number of positive responses. As noted, variables that take values only for positive responses generally have low rates of consistency between 2003 and 2005. Among these 30 variables, the highest percentage of cases with an “X” (or “2”) in 2005, after having an “X” in 2003, was 17 percent.
- **Unit feature variables:** Once again, the X-variables display a puzzling lack of consistency. Among those answering positively in 2003, the percent responding positively in 2005 was 11 percent for STEAM (steam system used as supplemental heating equipment), 14 percent for FLOT (heating equipment vented to outside), and 42 percent for FRPLI (fireplace has inserts). STEAM and a number of other variables related to supplemental heating were deleted from the 2007 AHS survey.
- **Testing variables:** Four variables relate to testing or inspection in the process of purchasing a house; these were TALWIR (aluminum wiring), TASB (asbestos), TRADON (radon), and TWATER (water). These variables, which were supposedly asked of all owner-occupied units, had between 9,000 and 10,000 cases with useable values. “Don’t know” or “refused” equaled 8 or 9 percent of the useable answers for the four testing variables: TALWIR, TASB, TRADON, and TWATER. The four testing variables had only moderately high levels of consistency given the nature of the questions; identical answers were obtained between 77 and 84 percent of the time for these questions. It is not apparent to the author that these variables are sufficiently important or accurate for continued inclusion in the AHS.

Important classes of variables that need improvement

There are two classes—variables related to the financing of owner-occupied housing and variables that identified government assisted housing—that produce information of fundamental importance in understanding how the housing market functions and how well it meets the needs of the population. However, many of the variables in these two classes have well-recognized shortcomings. HUD and the Census Bureau should take steps to improve these variables.

- **Mortgage-related variables:** With the demise of the Residential Finance Survey and the crisis in the mortgage market, the mortgage-related information in the AHS acquires a high level of importance. Unfortunately, many improvements are needed in this segment of the AHS.
 - The questions used to identify different mortgage products (fixed-rate mortgages, ARMs, GPMs) need to be updated to reflect current lending practices, and the variables need to be simplified so PUF users do not require special code to translate the variables into the mortgage product categories. Improving this component of the survey will require careful judgment, because many of the mortgage products popular in the last 5 years are now out of favor.
 - MORTIN (type of 1st mortgage) also needs to be improved. It relegates private mortgage insurance to “some other type” category.
 - A number of important variables related to mortgage finance have high rates of edits and allocations:
 - AMMORT – amount of 1st mortgage – was allocated for 20 percent of the cases.
 - VALUE – current market value of the unit – was allocated for 20 percent of the cases
 - INTW – interest rate of 1st mortgage, whole number – and INTF – interest rate of 1st mortgage, fraction – were allocated for 20 percent of the cases.
 - TERM – term of 1st mortgage – was allocated for 10 percent of the cases.
 - PMT – monthly payment on 1st mortgage – was edited for 13 percent of the cases.
 - MATBUY – got 1st mortgage in same year unit bought – was edited in 10 percent of the cases.
 - Consistency in responses across surveys is also a problem for some of the mortgage-related variables. Consistency was determined by comparing the value of a variable in 2003 with the value in 2005 for those sample cases where *the household was the same in both years and the unit is owner-occupied, with a mortgage reported as originated during the same year in both surveys.*

- MATBUY – got 1st mortgage in the same year bought: Different answer in 20 percent of cases.
- MORTIN – type of 1st mortgage: Different answer in 20 percent of the cases.
- REFI – mortgage a refinance of previous mortgage: Different answer in 21 percent of cases.
- TAXPMT – property taxes included in mortgage payment: Different answer in 14 percent of cases.
 - TERM – term of 1st mortgage: Answers differed by more than 3 years in 24 percent of the cases.
 - LOWINT – 1st mortgage refinanced to get lower interest: Among those who said “yes” in 2003, 56 percent said “yes” in 2005.
- The AHS does not collect information on the outstanding balances on mortgages other than home equity lines of credit. Finance analysts said the absence of this information is a major defect in the mortgage-related information provided by the AHS.
- **Assisted housing variables:** The variables used to identify assisted housing—RENEW, REPHA, SUBRNT, PROJ, VCHER, and APPLY—performed well in our analysis of (a) the number of cases with useable values and the prevalence of edits and allocation, and (b) the consistency of responses by the same household between 2003 and 2005. Unfortunately, the AHS has a recognized tendency to overestimate the number of assisted units, which implies that units identified as assisted are really not assisted. The National Academy of Sciences Panel that evaluated HUD’s research and evaluation programs recommended that HUD and the Census Bureau consider using address matches between HUD program records and the AHS sample to identify assisted units more accurately.

Variables that HUD and the Census Bureau should examine further for possible improvement or trimming

- **Utilities:** The use and costs of utilities are noteworthy issues in themselves and, in addition, they comprise an essential component of overall housing costs, which is, perhaps, the most policy-important of all the dimensions of housing. As shown in Section 7, the average costs of electricity, gas, fuel oil, other fuels, trash removal, and water and sewer services vary substantially both cross-sectionally and over time. The extent of the observed variation underscores the need for reliable information on utility costs.

Yet difficulties persist. Respondents frequently do not have good recall about utility expenses and even when recall is good, it can be affected by seasonal fluctuations in utility bills. The few studies that have been conducted regarding respondent error suggest that residents overestimate their utility costs. Over the 35-year history of the AHS, the Census Bureau has refined the way it asks about utility costs and how it processes the responses it receives. While these changes have improved the accuracy of the reported data, the results have not been tested against actual utility bills since 1987. HUD and the Census Bureau should consider updating this test.

In addition, HUD and the Census Bureau should review whether all the questions associated with utilities are needed. We have already recommended eliminating the BILL questions. Recent analysis funded by HUD has found inconsistent answers to the question regarding type of heating fuel. Revised wording or better edits may be needed for this and similar questions.

On March 19, 2009, the authors met with representatives of HUD and the Census Bureau at the Census Bureau to discuss a draft of this report. At that meeting, someone made a suggestion that we believe should be included in the recommendations of this report.

If HUD and the Census Bureau decide to eliminate any variables or groups of variables, they should consider testing whether the absence of the information provided by these variables will be missed. This test can be accomplished by phasing in the elimination. Specifically, after HUD and the Census Bureau make the decision, they should collect and code the variables for one additional survey but not include the variables on the public use file. If no one complains about the absence of the variables within the period between surveys, the variables should be eliminated from the questionnaire in the next survey.

4. Mortgage-Related Variables

In recent years, HUD has examined the usefulness of the information on mortgages because of the growing importance and complexity of housing finance. In February 2004, HUD published a study conducted by Abt Associates of Cambridge, Massachusetts. The study was entitled *Analysis of Housing Finance Issues Using the American Housing Survey (AHS)* and was prepared by Ken Lam and Bubul Kaul.³ This report examines the strengths and weaknesses of the mortgage data produced in the AHS from the perspective of the data's ability to shed light on the most common research and policy issues in housing finance. In November 2007, HUD released a study by Econometrica, entitled *The American Housing Survey and Non-Traditional Mortgage Products*, which was authored by Fred Eggers and Don Bradley.⁴ This second report examines the questions in the AHS on mortgage finance, with the objective of designing questions to obtain more information on the nontraditional mortgage products that have become so prevalent in housing finance. Both reports identified strengths and weaknesses of the AHS for studying issues involving housing finance. A third, HUD-funded report organizes the data on the financing of owner-occupied units from the five AHS surveys conducted between 1997 and 2005.⁵ This report showed researchers and policymakers in the housing finance area that the AHS contains valuable information on mortgages.

³ This study is available at the HUD USER Web site, www.HUDUSER.org, under *Publications*.

⁴ This study is available on the HUD USER Web site at <http://www.huduser.org/publications/polleg/ahsanalysis.html>.

⁵ Frederick J. Eggers, *Data on Home Mortgage Finance from the 1997, 1999, 2001, 2003, and 2005 American Housing Surveys*, prepared by Econometrica, Inc., April 2008. This study is available on the HUD USER Web site at http://www.huduser.org/intercept.asp?loc=/publications/pdf/ahs_mortgage.pdf.

Type of mortgage instrument

Both the Lam and Kaul and the Eggers and Bradley papers call attention to the weaknesses of the questions in the AHS on type of mortgage instrument. The AHS collects information related to mortgage instruments that are little used today, such as graduated payment mortgages, but does not collect information needed to identify the mortgage instruments that characterized the reckless lending period of 2005 and 2006, such as interest-only mortgages or option ARMs. In addition, users need special code to translate the AHS variables into the mortgage product categories reported in the AHS publications.

A number of the variables currently used by the AHS to identify type of mortgage product had low counts of cases with useable values. For example, BLOON (1st mortgage, last payment biggest) had only 53 sample cases with useable values.

Mortgage insurance

Both studies commissioned by HUD found that the AHS does not have good counts of mortgages with private mortgage insurance. In fact, the AHS variable MORTIN relegates private mortgage insurance to “some other type” category.

Mortgage debt

The AHS does not collect information on the outstanding balances on mortgages other than home equity lines of credit.

Small counts and extensive edits or allocations

The analysis of cases with useable values and the extent of allocations and edits identified several concerns among the mortgage-related variables.

- AMMORT – amount of 1st mortgage – was allocated for 20 percent of the cases.
- VALUE – current market value of the unit – was allocated for 20 percent of the cases.
- INTW – interest rate of 1st mortgage, whole number – and INTF – interest rate of 1st mortgage, fraction – were allocated for 20 percent of the cases.
- TERM – term of 1st mortgage – was allocated for 10 percent of the cases.
- PMT – monthly payment on 1st mortgage – was edited for 13 percent of the cases.
- MATBUY – got 1st mortgage in same year unit bought – was edited in 10 percent of the cases.
- AMTM – amount of other charges included in mortgage payment – had only 125 sample cases with cases with useable values.
- HELCN – number of home equity lines of credit – was reported for only 4,022 households. This count is consistent with the AHS estimate of the number of owner-occupied households with home equity lines of credit in 2005, 14.3 percent. The AHS and the Residential Finance Survey had roughly equal percentages of owner-occupied households with home equity lines of credit in 2001, 9.5 percent and 12.2 percent.

Consistency of responses: 2003–2005

For the mortgage variable, consistency was determined by comparing the value of a variable in 2003 with the value in 2005 *for those sample cases where the household was the same in both years and the unit is owner-occupied with a mortgage reported as originated during the same year in both surveys*. For some variables, the Census Bureau uses dependent interviewing, that is, it uses information obtained in previous surveys rather than re-asking a question. Among the mortgage variables, DWMPAY (amount of down payment) is an example of dependent interviewing.⁶

Among the mortgage-related variables there were some troubling examples of inconsistency.

- MATBUY – got 1st mortgage in the same year bought: Different answer in 20 percent of cases.
- MG – any mortgage on this property: Different answer in 13 percent of cases.⁷
- MORTIN – type of 1st mortgage: Different answer in 20 percent of the cases.
- REFI – mortgage a refinance of previous mortgage: Different answer in 21 percent of cases.
- TAXPMT – property taxes included in mortgage payment: Different answer in 14 percent of cases.
- TERM – term of 1st mortgage: Answers deferred by more than 3 years in 24 percent of the cases.⁸
- ARM – 1st mortgage changes due to interest rate: Among those who said “yes” in 2003, 38 percent said “yes” in 2005.
- FIXED – 1st mortgage changes for taxes/insurance: Among those who said “yes” in 2003, 27 percent said “yes” in 2005.
- LOWINT – 1st mortgage refinanced to get lower interest: Among those who said “yes” in 2003, 56 percent said “yes” in 2005.

A number of important mortgage-related variables were consistent between 2003 and 2005. For PMT (monthly payment on 1st mortgage), the 2005 value was within 20 percent of 2003 value for 67 percent of the cases. For NUMMOR (number of mortgages including home equity loans), the 2005 value was within one digit of the 2003 value for 98 percent of the cases.

5. Neighborhood Conditions

The authors decided to focus on neighborhood variables because, to their limited knowledge, these variables have not been used extensively. Nine of the 89 articles reviewed appeared to have made use of the neighborhood variables available in the AHS PUF. Some of the papers have a specific neighborhood focus, while others use neighborhood information for control purposes. While neighborhood conditions are important in explaining many housing market and

⁶ The answers to DWNPAY differed in only 7 out of 7,129 cases between 2003 and 2005.

⁷ We selected the cases for comparison on the basis of MG =1 in 2005; it is possible that there was no primary mortgage on the unit in 2003.

⁸ The consistence test assumed that the term would be reduced by 2 years between 2003 and 2005.

related behaviors, the AHS neighborhood variables are not ideal for these purposes. As the AHS codebook notes: “neighborhood conditions are based on subjective assessments of the respondent rather than on objective evaluations of the neighborhood.... Furthermore, neighborhood is not defined to the respondent; it is whatever the respondent considers his/her neighborhood to be.”

The neighborhood variables contained in the PUF generally have high response rates. Most of the variables have cases with useable values for almost all units in the sample. The major exceptions are variables that involved followup questions, such as “is the noise/traffic problem so bad as to make you want to move?” (STRNC). Some variables are asked of only some units. For example, “whether there is a school close” (SCHCLS) is asked of approximately 12,000 units; the codebook does not contain information on “out of universe” for this variable. Of the 50 neighborhood variables analyzed, only one variable—shared plumbing facilities (SHARPF)—is subject to edits or allocations. Only one variable—SCH (is the neighborhood public elementary school satisfactory?)—had a worrisome level (14.5 percent) of “don’t know” answers or refusals to answer; almost all of these answers were “don’t know.”

The comparison of answers given by the same household in 2003 and 2005 to the neighborhood questions found a general level of consistency when the coding distinguished between positive and negative answers. There were some exceptions:

- COMMRECR: community recreation facilities available; 73 percent gave the same answer in both years.
- COMSERV: community services provided, 79 percent.
- EGREEN: open space available within half-block, 72 percent.
- ESFD: single-family homes within half-block, 87 percent.
- ETRANS: railroad/airport/four-lane highway within half-block, 87 percent.
- SCHCLS: neighborhood public elementary school within 1 mile, 79 percent.

When the variables were coded to identify only those cases that had responded positively to a question, then consistency was surprisingly bad. Because the questions generally dealt with neighborhood problems, a “positive” response meant that there was a problem. For example, BADPER (people in neighborhood are bothersome) is coded either “X,” indicating the presence of the problem, or left blank.⁹ Normally, one would not expect much consistency between variables identifying specific problems, but the observed level of consistency is disturbing. Here are some examples:

- BADPER: 10 percent of households who cited this problem in 2003 did so again in 2005.
- BADSERV: poor city/county services are bothersome, 11 percent.
- LITTER: litter in neighborhood is bothersome, 15 percent.
- NOISE: noise in neighborhood is bothersome, 12 percent.

⁹ “B” (not applicable) or “R” (refused) are also possible codes.

Only seven neighborhood variables are coded in this manner; they all showed very low consistency. The authors wonder about the usefulness of these “X” variables in the neighborhood context. There are two alternatives: The inconsistency represents a problem with the variable, or the inconsistency reflects real changes between surveys. Even in the latter case, one can question the usefulness of the information provided. If we assume that the 2003 and 2005 responses are both correct, then “X” variable does not appear to provide a useful description of conditions in the neighborhood, since the problem appears to be ephemeral.

6. Assisted Housing¹⁰

For research and policy analysis, it is important to be able to determine whether the government assists households with the costs associated with living in their units. Because housing assistance is a central component of federal housing policy, housing analysts would like to be able to identify and analyze assisted units within the AHS sample. In addition, tenants often behave differently in subsidized units; for example, they may be less likely to move if moving involves losing the subsidy. Therefore, any analysis of housing behavior using the AHS would benefit from being able to distinguish assisted from unassisted units.

The AHS collects information from households that analysts can use to label units as assisted or unassisted. The AHS has a recognized tendency to overestimate the number of assisted units, which implies that units identified as assisted are really not assisted.¹¹ This section will not discuss this issue; instead it will focus on the quality and consistency of the data collected related to assistance. We do suggest, however, that HUD and the Census Bureau follow the recommendation of The National Academy of Sciences Panel to consider using address matches between HUD program records and the AHS sample to identify assisted units more accurately.

HUD’s Housing Affordability Data System (HADS) defines assisted housing using six variables. The HADS’ definition is:

In the 1997 and later datasets, a unit is assisted if:

- RENEW = '1' AND (REPHA = '1' OR REPHA = '2') [Income reporting needed for lease renewal and this is reported to the landlord, building manager, or housing authority]
- OR SUBRNT = '1' [Household receiving a subsidy]
- OR PROJ = '1' [Unit is public housing]
- OR VCHER='1' [Household has a voucher]
- OR APPLY='1' [Applied to management to get public housing]

The AHS asks these questions only of households living in rented units. RENEW and SUBRNT are asked of most households in rental units. Both variables had a large number of cases with useable values in 2005 (13,585 and 15,015, respectively), and the percent of answers that were “don’t knows” or refusals was 1 percent or less. The other four variables are followup questions.

¹⁰ See Appendix for a discussion of the changes to the instrument in the 2007 AHS.

¹¹ See page 11 of the HADS documentation for references related to this overestimation.

PROJ had 4,242 usable answers; there were 1,452 edits associated with this variable. VCHER had 669 useable answers; the percentage of “don’t knows” or refusals was 5.9 percent.

Consistency in answers by the same household between 2003 and 2005 was very good: 97 percent gave the same answer in both years for RENEW, 99 percent for SUBRNT, 98 percent for REPHA, VCHER, and APPLY. The Census Bureau uses dependent interviewing for PROJ; that is, the Census Bureau uses the answer from the previous year.

7. Housing Costs, Including Utilities

From both analytical and policy perspectives, data on housing costs are among the most important information collected by the AHS. This section focuses on RENT (the amount of rent paid) and utility costs. The section on mortgage-related variables discusses PMT (monthly payment) for mortgage and related costs.

The AHS collects information both on RENT and the frequency of payments (FRENT). The Census Bureau allocated values to RENT and FRENT in 11.4 percent of the useable cases and edited RENT in an additional 4.6 percent. In comparing RENT in 2003 with its value in 2005, we assumed 15 percent inflation over the period; it turns out that the sum of rent plus trash plus water grew by 7 percent over the period. Under our assumption, 73 percent of the 2005 values were within 20 percent of the inflated 2003 value.

The utility variables deserved special attention due to long-standing concern about the reliability of these data. The AHS devotes considerable effort to collecting information on the use of and the costs of utilities. For each utility, the Census Bureau asks respondents:

- whether they use the utility;
- if the respondent is a renter, whether he or she pays for the utility or whether it is included in rent;
- if the respondent pays for the utility, whether its costs are billed with another utility; and
- the costs of the utility
 - For electricity and gas, respondents are asked to provide bills for January, April, August, and December.

This battery of questions was designed to ensure that all utility costs are identified and associated with the appropriate utility. In addition, the Census Bureau uses regression formulas based on the Residential Energy Consumption Survey to adjust whatever cost data are provided for electricity and gas.

Despite the level of detail gathered and the care taken by the Census Bureau, there remain concerns about the reliability of the information provided by respondents. A current study funded by HUD looked at data on utility usage collected from a subset of households most likely to provide reliable answers, namely, occupants of three-bedroom homes in single-unit detached structures with resident-paid utilities. By comparing AHS data with Residential Energy Consumption Survey (RECS) data, the researchers conclude the respondents sometimes misidentify the type of heating equipment they have.

The variables related to utility costs are difficult to assess. Table 1 presents information from our cases with useable values and consistency analysis.

Table 1: Edit and Allocation Rates and Consistency for Utility Cost Variables

Variable	Definition	Number of cases with values	Percent of cases with useable values edited	Percent of cases with useable values allocated	Percent of 2005 responses within 10% of 2003 responses adjusted for change in means	Percent of 2005 responses within 30% of 2003 responses adjusted for change in means
AMTE	Average monthly electric	41,013	0.0%	100.8%	23.9%	56.5%
AMTG	Average monthly gas	24,928	0.0%	108.1%	24.2%	60.0%
AMTO	Annual fuel oil	3,247	1.1%	33.2%	24.4%	62.2%
AMTF	Annual other fuels	2,352	56.3%	91.9%	15.8%	35.4%
AMTT	Annual trash	18,508	5.7%	20.6%	33.3%	55.6%
AMTW	Annual water	23,661	6.4%	20.2%	22.4%	48.0%

The average monthly cost of electricity is the most important utility variable in the sense that 41,013 units report this cost, while only 2,352 units report annual costs for other fuels. The number of units reporting costs for trash and water is relatively low, because the costs of water and trash are frequently included in rent. The allocation rates are quite high. The number of cases edited or allocated exceeds the number of cases with useable values for electricity, gas, and other fuels. The Census Bureau uses hot deck allocation for all of these variables and, in addition, it uses regression to allocate electricity and gas. The variables used to indicate an edit or allocation for gas and electricity take separate values for hot deck and regression allocation.

To assess the consistency with which the AHS reports utility costs, we compared the utility costs in 2003 to utility costs in 2005, unit-by-unit, for all units occupied by the same household in both years. To obtain an accurate measure, we deflated the 2005 costs of each utility by the percent change between 2003 and 2005 in the mean cost for owner-occupied units of that utility.¹² Household expenditures on electricity, gas, and other utilities can fluctuate for reasons other than the cost of the underlying service. It is difficult to judge whether the observed level of consistency between 2003 and 2005 responses is reasonable or not. In general, we believe that the percentage of cases within 30 percent should be greater than the percentages observed in Table 1.

¹² The adjustment factors were 4 percent for electricity, 23 percent for gas, 21 percent for fuel oil, 1 percent for other fuels, 1 percent for trash removal, and 7 percent for water and sewer.

Because of the inherent importance of information on utility costs, we investigated how much variation there is across time and across major subgroups of the population in reported utility costs. Table 2 presents the results of that analysis.

Table 2: Variation in Utility Costs Across Time and Across Populations

Averages calculated only for units with non-zero expenditures	Average monthly electric	Average monthly gas	Annual fuel oil	Annual other fuels	Annual trash	Annual water
1999	\$71.12	\$50.38	\$729.31	\$217.64	\$205.81	\$370.57
2005	\$77.96	\$77.98	\$1,319.81	\$298.70	\$258.12	\$445.99
In 2005						
Northeast	\$78.08	\$105.54	\$1,522.85	\$477.07	\$254.11	\$457.57
Midwest	\$68.48	\$92.53	\$943.29	\$240.59	\$216.73	\$406.81
South	\$90.49	\$68.43	\$912.00	\$244.27	\$244.30	\$417.32
West	\$66.60	\$53.18	\$789.05	\$243.39	\$314.94	\$535.66
Central city of MSA	\$72.06	\$78.22	\$1,348.80	\$198.19	\$301.86	\$461.29
Inside MSA, but not in central city - urban	\$78.03	\$81.03	\$1,533.54	\$229.36	\$273.91	\$464.01
Inside MSA, but not in central city - rural	\$89.51	\$81.80	\$1,334.51	\$333.08	\$241.11	\$440.76
Outside MSA, urban	\$70.88	\$73.53	\$1,114.52	\$310.32	\$228.57	\$435.91
Outside MSA, rural	\$82.18	\$68.56	\$950.00	\$375.68	\$210.09	\$374.76
Non-Hispanic, white only	\$78.77	\$78.96	\$1,318.48	\$305.92	\$247.11	\$441.64
All other	\$75.75	\$75.33	\$1,322.41	\$263.40	\$301.99	\$459.06
Less than or equal median household income	\$69.06	\$73.82	\$1,161.60	\$340.36	\$237.68	\$393.26
Above median household income	\$86.51	\$81.41	\$1,436.76	\$270.92	\$271.64	\$484.44

Table 2 reports substantial variation across regions, within metropolitan areas and between metropolitan areas and non-metropolitan areas, and by household income across all utilities.¹³ Major variation between households with non-Hispanic, white only householders and all other households is confined to the annual costs for trash removal. The change in mean costs between 1999 and 2005 varied greatly across the various utility items.

The AHS includes 30 variables that indicate whether a given utility is billed in conjunction with another utility. For example, BILLGE tells whether gas is billed with electricity. These variables take the value of “X” (or sometimes “2”) if the answer is yes; “B” (not applicable) and “blank” (not reported) are the only other values allowed. BILLW (water billed with other utility) has the most number of positive responses, 1,525. The average number of “yesses” is 173. The Census Bureau does not allocate values to any of these variables, but edits are frequent, ranging from 43,358 cases with edits to 30 cases. For every variable, the number of edits exceeds the number of positive responses. As noted earlier, variables that take values only for positive responses generally have low rates of consistency between 2003 and 2005. Among these 30

¹³ Because these means are calculated using only those households with non-zero costs, it would be inappropriate to add the various mean costs together.

variables, the highest percentage of cases with an “X” (or “2”) in 2005, after having an “X” in 2003, was 17 percent. Unlike the neighborhood “X” variables, there is no reason to suppose that the facts underlying response to the BILL questions would have changed in this 2-year period.

The AHS also provides six variables to indicate whether the occupants of a unit pay separately for each of the various utilities. For example, BUYE takes the value of “1” if electricity is not used, “2” if included in rent or other fees, or “3” if obtained free. These variables are recodes added to the file to provide the data users variables just like the ones they received prior to the 1997 AHS national file. They take “B” (not applicable) values under a variety of circumstances, including if the reported cost falls within an allowable range. Since these variables are recodes, they impose no burden on respondents and cost little to produce. However, our tests do raise some questions about these variables. Allocations are frequent among these variables, ranging from 6 percent to 246 percent (BUYE). We did not run consistency tests on these variables.

The set of questions dealing with utilities is designed to answer a number of different questions. Of primary policy importance are what utilities are used and how much does each utility cost the household. We have no independent source of information to judge how well the reported costs correspond to actual costs. Presumably, the three-step allocation procedure explained on page C-25 of the 2007 AHS National Report produces reasonable information. However, the accuracy of the procedures has not been tested against actual utility bills since 1987. HUD and the Census Bureau should consider updating this test.

The consistency checks show greater percentages of cases more than 30-percent different from the previous response than we believe are reasonable. The data on variability indicate that year-to-year difference and group-to-group differences are important and cannot be ignored.

Information on how utilities are used and how they are billed are important only to the extent to which they contribute to the answers of the prime questions. The BILL questions produce answers that are very inconsistent across years, and combined billing does not seem prevalent enough to warrant this effort. The recent HUD-sponsored research also raises questions about the information collected on how utilities are used.

8. Housing Features and Amenities

The AHS codebook chapter labeled “Unit” deals with variables related to the size, features, and amenities of a unit and its ownership.

An analysis of 94 variables in this group found that the count of cases with useable values (responses other than “refused” or “don’t know”) was consistent with the information being sought. The proportion of “don’t knows” or “refused” was high for two variables: 20 percent for both MOPERM (number of months since occupied as permanent home) and SRENTV (number of nights vacant/URE unit was rented). While “don’t knows” and “refused” were not as high, they were substantial (9.8 percent) for UNITSF (square footage of unit).

There were a few variables for which the number of cases with useable values was so low that one wondered about the value of collecting the information. These include STEAM (steam

system used as supplemental heating equipment), for which 210 cases replied positively; SINHV (inherited this vacant/URE unit), for which 179 cases responded positively; and SSELV (want to sell vacant/URE unit, not sold yet), for which 34 cases responded positively. STEAM and a number of other variables related to supplemental heating were deleted from the 2007 AHS survey.

ELEV (working passenger elevator) had a high percent of edits (63 percent). Allocations were high for LOT (square footage of lot)—17.6 percent; OVEN (unit has a working microwave oven)—11.9 percent; and BURNER (unit has working built-in cooking burners)—10.4 percent. The last two variables are recorded only for units without a working stove, and both are recorded using dependent interviewing.

Consistency in responses between the 2003 and 2005 surveys was generally good for these variables. A number of the variables, such as UNITSF, LOT (lot size), and WHNGET (year unit bought/obtained/received), are collected using dependent interviewing; therefore, one should expect very high consistency. CONDO (unit is condominium or cooperative) was answered the same in 97 percent of the cases involving the same household. This is an important variable and one not subject to dependent interviewing. Two variables had high rates of consistency between surveys, but are so specific that one wonders why even higher consistency was not obtained: DINING (number of dining rooms in unit) was answered identically in 78 percent of the cases, and WFPROP (unit is waterfront property) was answered identically in 92 percent of the cases.

Once again, the X-variables display a puzzling lack of consistency. Among those answering positively in 2003, the percent responding positively in 2005 was 11 percent for STEAM, 14 percent for FLOT (heating equipment vented to outside), and 42 percent for FRPLI (fireplace has inserts).

9. Housing Quality

Data in the Unit Quality section include a measure of overall quality of the unit, residents' satisfaction with the unit, deficiencies outside and inside the unit (e.g., leaks, holes, cracks, blown fuses, equipment breakdowns) and in the common halls (such as nonworking light fixtures or loose railings), and maintenance and repair information. Most of the variables relate to deficiencies; variables related to amenities and other positive features are usually found in the Unit section of the codebook. We analyzed 69 variables from this section.

Many of the quality variables allow multiple answers or, if a yes/no question, both the “yes” and the “no” responses are coded. The variables related to both owner-occupied and renter-occupied units generally have a large number of useable answers; 22 variables have useable answers for 40,000-plus cases. Some variables relate only to rental units; 10 variables in this group had useable answers for 10,000-plus cases. Four variables that relate to testing or inspection in the process of purchasing a house and that were supposedly asked of all owner-occupied units had 9,000 and 10,000 cases with useable values; these were TALWIR (aluminum wiring), TASB (asbestos), TRADON (radon), and TWATER (water). The variables that had fewer than 5,000 cases with useable values were generally X-variables or variables that were asked only if a

specific problem occurred, such as IFCOLD (Was that because the MAIN heating equipment broke down?) or NUMTLT (number of toilet breakdowns, 6 hours or more).

The variables related to the presence of rodents had a high level of edits and allocations. The cases with useable values for M3ROD (rodents seen in unit recently) were edited in 39 percent of the cases; the 121 cases with useable values for NOTSUR (not sure if rodents were rats or mice) were allocated in 26 percent of the cases. Allocations accounted for 9 percent of the cases with useable values for RATS (rats seen in unit recently).¹⁴ “Don’t knows” or “refused” equaled 8 or 9 percent of the useable answers for the four testing variables: TALWIR, TASB, TRADON, and TWATER. Otherwise, incomplete responses and edits or allocations were not important for the quality variables.

As noted earlier, most of the variables in the Unit Quality section identify deficiencies. If one assumes that owners or landlords eliminate or alleviate deficiencies over a 2-year period, then one would not expect consistent responses to most of the Unit Quality variables between the 2003 and 2005 surveys. Actually, for the variables that allow yes/no, this expectation is generally not correct. For example, results for HOLES (holes in floor) show identical answers in 2003 and 2005 for the same households in 99 percent of the cases. This high level of consistency probably relates to the general high-quality level of the stock. In 2005, only 0.8 percent of occupied units had holes in the floor (HOLES). If the same percentage applied in 2003 and the distribution of units with holes was random in both years, one would expect identical answers to HOLES in 98 percent of the cases. So consistency results from low level of deficiencies rather than from failures to make improvements.

Even for variables with multiple answers, the level of consistency between 2003 and 2005 is high. For example, MAJR1 (landlord starts major maintenance/repairs soon enough) allows five coded responses, and identical responses were recorded in 64 percent of the cases. The variables that counted instances of problems, such as NUMTLT or NUMBLOW (number of times fuses blown or breakers tripped) were also reasonably consistent, but as noted, the number of responses was small. The answers to NUMTLT differed by no more than one in 67 percent of the cases.

The four testing variables (TALWIR, TASB, TRADON, and TWATER) had only moderately high levels of consistency given the nature of the questions; identical answers were obtained between 77 and 84 percent of the time for these questions. The X-variables displayed the usual low level of consistency. For some of the variables, such as MICE (mice seen in unit recently), the low level of consistency may reflect changed conditions between 2003 and 2005. But this rationale may not apply to all the X-variables in this group. For example, among the 176 units where households reported they were cold in 2003 because of inadequate insulation (WHYCD3), only 18 (10 percent) of the same households gave the same explanation in 2005. However, the WHYCD3 question was considered inappropriate for 140 of the 176 households in 2005.

The overall measures of quality were consistent between 2003 and 2005. Units that were occupied by the same households in 2003 and 2005 gave the unit the same overall satisfaction

¹⁴ At the March 19, 2009, meeting, Census Bureau staff explained that the relationship between the rodent variables accounted for the high number of edits and allocations.

rating—on a scale of 1 to 10—in 41 percent of the cases. The overall rating in 2005 was the same or within one digit of the 2003 rating in 73 percent of the cases and was the same or within two digits in 90 percent of the cases. ZADEQ classifies units into three categories: adequate, moderately inadequate, or severely inadequate. Units with the same households had identical ZADEQ classifications in 2003 and 2005 in 94 percent of the cases. Again, this high level of consistency probably relates to the general high-quality level of the stock. In 2005, 94 percent of the occupied stock was adequate; 4 percent, moderately inadequate; and 2 percent, severely inadequate. If the same proportion existed in 2003 and if the distribution was purely random in both years, we would expect to find identical classifications in 89 percent of the cases.

10. Household Composition

Our limited literature review confirmed the importance of basic information on households and householders. Age, race, gender, and general demographics were used in a large number of the cited works. Family or household size, the presence or absence of children, and immigrant status were used in several studies. Surprisingly, education and employment status appeared as variables in only a couple of studies. Their limited usage probably says more about the limitations of our literature review than about the importance of these variables.

The variables grouped in the Household Composition section of the codebook describe the household, the householder, and other household members. In addition, there is a subset of variables that describe the relationship among household members. Some variables related to income are also listed under Household Composition.

We will discuss the household composition variables in two stages. First, we will look across all the variables, focusing on the number of times the variables are used, the frequency of edits and allocations, and the consistency of responses in successive surveys. In this analysis, we will only look at household members 1 and 2 for variables that are reported for all household members. In the second stage, we will look at all household members for these variables.

In the first stage of the analysis, we looked at 64 variables. For 25 variables, there were cases with useable values for over 40,000 cases. For another 9 variables, there were 30,000 useable cases. Eight of these variables describe the characteristics of the second household member; the counts of cases with useable values conform to the number of households with more than one member. Fourteen variables had fewer than 10,000 cases with useable values and 3 had fewer than 1,000 cases with useable values. In every case, the subject matter of the question—or in some cases, the fact that X-variables do not record “no” answers—accounts for the small number of cases with useable values.

“Don’t knows” are a significant issue for SNITEV (number of nights owner spends at vacant/URE unit) and WHHRW1 (number of hours worked at home [wage & salary] for the householder). For SNITEV, there were 609 “don’t knows,” 23 refusals, and 1,938 cases with useable values; for WHHRW1, there were 223 “don’t knows,” 16 refusals, 78 edits, and 2,916 cases with useable values. Edits were very important for HERE2 (nonrelative is present, second household member)—29.6 percent of the cases with useable values—and SPOS1 (line number of spouse of householder)—25.5 percent of the cases. Allocations exceeded 5 percent of the cases

with useable values for QIDRI (family receives interest/rental/royalty/estate income), QSELF (family receives self-employment income), QOTALM (family receives VA/unemployment insurance/child support/alimony/other), QRETIR (family receives retirement or survivor pension), and QSSI (family receives supplemental security income).

HERE2 was consistent between 2003 and 2005 in only 65 percent of the same households. Otherwise, consistency between 2003 and 2005 was good. In many cases, consistency was almost automatic because of dependent interviewing.

We decided to look at key demographic characteristics of all household members, on the assumption that missing values and edits and allocations might be more likely for some of these variables, especially for members of large households. Here are some conclusions:

1. The AHS provides full demographic information on all household members. There are no missing values for any of these variables for any of the persons. (This conclusion excludes SPOS, because we do not have a count of spouses to check the numbers against. It also presumes that variables with “B” [not applicable] values are correctly identified for GRAD and MAR.)
2. Generally, the rates of edits or allocations are low. SPOS, MOVN, and SAL are notable exceptions. Approximately 5 percent of the RACE values were edited or allocated. (We used the number of persons as the denominator in all cases. For GRAD and MAR, this includes “Bs”.)
3. For some variables, values are only edited, never allocated. These include REL, SPAN, and SPOS. This seems surprising for SPAN. The AGE and GRAD variables are only allocated, never edited.
4. We would have expected the rate of edits or allocations to increase as one obtained information on each additional household member. This pattern exists for some variables, such as SEX, SPAN, and RACE, except that there are generally no edits or allocations for 14th through 16th household members. The absence of edits and allocations for these members is surprising.

11. Recent Movers

Data on recent movers are shown for households where the householder or a member of the household moved into the present unit during the 24 months prior to the interview. Questions in this section ask about the reasons for moving and for choosing the current unit. Questions about the previous residence are also asked of every recent mover in the household, while the questions about the reasons for moving and choice of unit are asked only of mover respondents. In 2005, 13,558 sample households were classified as recent movers.

We analyzed all 93 recent mover variables in the 2005 AHS. Some variables are asked of all households, such as MOVE1-16 (year person moved into unit). Five variables had over 40,000 useable cases. Of these, MOVE1 (year householder moved in) was edited in roughly 1 percent

of the cases and allocated in roughly 1 percent. MOVMI (month householder moved in) was edited in roughly 3 percent of the cases and allocated in 13.5 percent. Among the variables reported for only recent mover households, ALMV (all movers lived in same previous residence) was allocated in 43 percent of the cases with useable values; WHYTOH (main reason this unit was chosen), WHYTON (main reason this neighborhood was chosen), and WHYMOVE (main reason moved) had edits that equaled approximately 70 to 75 percent of the cases with useable values. Relating edits to cases with useable values is difficult because sometimes edits far exceed the number of cases with useable values; for example, edits equaled over 250 percent of cases with useable values for MV1 (recent mover group 1) and MVG2 (recent mover group 2).

Some variables had few cases with useable values because they were X-variables where only “yes” is recorded, and the variables dealt with limited situations, such as WMONHH (moved to establish own household), which had 1,761 positive responses.

We did not test the consistency of the recent mover variables because a household identified as a recent mover household in the 2003 survey would not be asked the recent mover questions in the 2005 survey, and a household identified as a recent mover household in 2005 would by definition not be the same household as the household interviewed in 2003.

12. Commuting

Journey-to-work is an important research issue independent of housing. Still, location with respect to places of work is a factor used in all explanations of how the housing market works. The AHS collects information on the number of household members who travel to work, the times and distances involved, and the modes of transportation used. It also has questions on carpooling, departure times, and car ownership. The recent mover segment of the questionnaire contains information on working at home.

The journey-to-work concept captures how members of the household commute to and from work. This commuting information is at the person level. Information on the number of vehicles is at the household level. In our limited search of the literature, we found five articles that made use of the AHS’s journey-to-work data.

There are only 12 variables in the commuting section, but most are collected for multiple household members. We looked at the variables only for household members #1 and #2. Of the 20 variables we examined, allocations were an issue only for distance-to-work variables DISTJ1 (8.3 percent) and DISTJ2 (8.1 percent). Seventeen percent of all respondents replied “don’t know” to TRN (is the public transportation satisfactory?). Unlike most “don’t knows,” this answer actually provides useful information.

The distance-to-work and time-to-work variables for household members 1 and 2 displayed a surprising lack of consistency between surveys. The answers in 2005 were within 30 percent of the 2003 answers only approximately 50 percent of the time. Because of this lack of consistency, we decided to look at variation in responses across time and across different geographies and different segments of the population. If there were little variation, we might

consistency, we decided to look at variation in responses across time and across different geographies and different segments of the population. If there were little variation, we might have considered recommending that HUD and the Census Bureau eliminate this segment or scale it back significantly.

Table 3 shows how commuting time and distance varied between 1999 and 2005 and, within 2005, across regions, metropolitan and nonmetropolitan segments, and by race and ethnicity and by income. There was a surprising amount of variation, particularly between households with non-Hispanic, white-only householders and all other householders. Given the level of variation and the theoretical and empirical connections between housing and location, we decided that journey-to-work remains an important part of the survey. We do suggest that HUD and the Census Bureau investigate the lack of consistency between the distance-to-work and time-to-work variables between the same households in different years and determine where the questions need improvement.

Table 3: Variation in Commuting Across Time and Across Populations

AHS variable	DISTJ1 In miles	TIMEJ1 In minutes
	Distance to work for householder	Commute time for householder
1999	40.05	49.10
2005	49.45	58.39
In 2005		
Northeast	44.90	56.65
Midwest	47.03	54.53
South	44.67	52.92
West	63.06	72.34
Central city of MSA	39.81	50.23
Inside MSA, but not in central city - urban	49.71	59.44
Inside MSA, but not in central city - rural	63.01	71.20
Outside MSA, urban	37.88	44.46
Outside MSA, rural	62.90	69.02
Non-Hispanic, white only	57.38	65.49
All other	30.19	41.13
Less than or equal median household income	42.31	51.04
Above median household income	53.69	62.74

13. Mobile Homes

Information was collected in the 2005 survey on only seven variables from this section. In the 2005 survey, there were approximately 3,000 mobile homes in the sample. Over 2,000 cases with useable values were obtained for all the variables except MHSTAY (mobile home to stay or be moved)—68 cases with useable values—which was apparently asked of only a few

mobile home). Consistency appears to be an issue only for TPARK (number of mobile homes in group where “1” = one mobile home and “2” = two or more mobile homes). While we normally consider 87 percent consistency good, 87 percent seems somewhat inconsistent, given the narrow choice of responses.

14. Lead Based Paint

Information was collected in the 2005 survey on only four variables from this section. BIGP (area of peeling paint larger than 8"x11") was asked of all households; the other three variables were asked of owner-occupants. The number of cases with useable values for these three variables was approximately 9,300. “Don’t knows” ranged from 700 to 800 and along with refusals constituted 8.4 to 9.5 percent of responses. Answers to BIGP were identical in 97 percent of the responses from the same households in 2003 and 2005. The other three variables had identical responses for 83 to 84 percent of the same households.

15. Upgrades and Remodeling

Information was collected in the 2005 survey on 16 variables from this section, one of which—RAS (type of alteration/repair)—accepts up to 16 answers drawn from 74 categories.

Four variables describing the types of changes made to an apartment had between 228 and 311 cases with useable values. The codebook does not provide “out of universe” information for these variables. With one exception, the remaining variables have between 9,200 and 29,600 useable cases. The exception—USFCAM (amount of square foot change in unit)—has 522 cases with useable values. This question is asked only if there is a change in square footage.

USFCHG (change in square footage of unit) has edits that equal 26 percent of the cases with useable values. “Don’t knows” were important only for responses related to costs, where they represented approximately 7 percent of the responses.

We did not perform consistency analysis on the variables in this section, because one would anticipate that upgrades and remodeling would change from survey to survey.

Appendix: Assisted Housing Items in 2007 AHS

Note: This appendix was added by HUD after the contractor submitted its final report.

Section 6 of this report (page 11) discusses the variables related to assisted housing as they appeared in the 2005 AHS instrument. The assisted housing questions were substantially revised in the 2007 survey. Because analysis of the changed variables was beyond the scope of Econometrica’s contract with HUD, this appendix has been added by the Department.

The changes to the AHS instrument were based on the findings of a study that HUD commissioned from ORC Macro, the results of which were published as *Improving Housing Subsidy Surveys: Data Collection Techniques for Identifying the Housing Subsidy Status of Survey Respondents* (2005, <http://www.huduser.org/Publications/pdf/datacollect.pdf>). That study used cognitive interviewing techniques to attempt to improve the assisted housing questions, because of the very problems that Econometrica identified in the current report.¹⁶ As a result, the AHS substantially changed the wording of the assisted housing questions in the 2007 survey instrument. The results of the changes can be seen in the table below, which presents estimates of assisted households in 2005 and 2007. The total number of assisted households, according to HUD records, remained essentially unchanged over the period, at approximately 4.3 million. Thus, the presumed overcount in the 2007 AHS is substantially smaller than in 2005. Moreover, the count of assisted households with incomes below 50 percent of area median income—which is the range in which most such households should be found—is approximately 4.3 million in 2007, roughly matching the estimate from administrative data. While the assisted housing questions in the AHS are not yet perfect (and probably never will be), the questions in the 2007 survey are a marked improvement over those of the 2005 survey discussed in section 6.

Assisted Households Identified in the American Housing Survey

<u>Year</u>	<u>Income as percent of Area Median</u>					<u>All</u>
	<u>0-30</u>	<u>30-50</u>	<u>50-80</u>	<u>80-120</u>	<u>≥ 120</u>	
2005	3,289	1,262	992	555	450	6,547
2007	3,350	1,016	468	166	161	5,161

¹⁶ The specific wording of the instrument items can be found in the instrument item booklets for the respective surveys, which can be downloaded from HUD USER, <http://www.huduser.org/datasets/ahs.html>. To find the beginning of the assisted housing sections, search the documents for the keyword RENEW, in upper case.