



HUD-Assisted Renter Oversample and Administrative Data Matching 2011 and 2013

LAST UPDATED: APRIL 2019

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
U.S. CENSUS BUREAU
U.S. DEPARTMENT OF COMMERCE





Contents

Purpose	1
What Users Need to Know	1
1. Historical Overview.....	2
1.1. Pre-2011 Administrative Data Matching	2
1.2. The 2011 HUD-Assisted Oversample and Integration Issues.....	2
1.3. The 2011 HUD-Assisted Administrative Data Matching and Matching Issues	3
1.4. Efforts to Address the Issues	3
2. HUD-Assisted Oversample Process	3
2.1. HUD-Assisted Oversample Frame Development	4
2.2. HUD-Assisted Oversample Selection.....	4
2.3. HUD-Assisted Oversample Integration and Weighting.....	5
2.4. Results and Discussion	6
2.5. Conclusion.....	7
3. HUD-Assisted Administrative Data Matching Process	7
3.1. Background.....	8
3.2. Addressing the Issue.....	8
3.3. Using the Administrative Data Match to Verify the Status of the HUD-Assisted Oversample	9
3.4. The Results.....	9
3.5. Removal of <i>HUDSAMP</i> from the 2011 and 2013 Public Use Files	10
4. HUD-Assisted Public Use File Weight <i>SUBWGT</i>	10
4.1. Development of <i>SUBWGT</i>	10
4.2. Use of <i>HUDADMIN</i> and <i>SUBWGT</i>	11
Recommendation 2.....	12
Appendix A.....	13
A.1. Comparison of Original and Revised Oversample Integration Method.....	13
References.....	15



List of Exhibits

Exhibit 2.1.	Summary of the HUD-Assisted Oversample Frame.....	4
Exhibit 2.2.	Initial and Final HUD-Assisted Oversample Frame	5
Exhibit 2.3.	Comparing the Oversample Frame With 2011 National Longitudinal Sample Estimates	6
Exhibit 2.4.	Comparing the Oversample Frame With 2013 AHS National longitudinal sample Estimates	7
Exhibit 3.1.	2011 AHS HUD-Assisted Totals Using WGT90GEO	9
Exhibit 3.2.	2013 AHS HUD-Assisted Totals Using WGT90GEO	10
Exhibit 4.1.	Comparison of 2011 AHS National Longitudinal Sample PUF HUD-Assisted Totals Using WGT90GEO and SUBWGT	11
Exhibit 4.2.	Comparison of 2013 AHS National Longitudinal Sample PUF HUD-Assisted Totals Using WGT90GEO and SUBWGT	11
Exhibit 4.3.	Example of When to Use WGT90GEO	12
Exhibit 4.4.	Example of When to Use SUBWGT	12
Exhibit A.1.	Comparison of Base Weighted HUD-Assisted Unit Counts for Original and Revised Releases	13
Exhibit A.2.	Size of Frame and AHS National Longitudinal HUD-Assisted Sample	13



Purpose

The purposes of this guide are to:

Describe the revised process for integrating an oversample of U.S. Department of Housing and Urban Development (HUD)-assisted housing units into the American Housing Survey national longitudinal sample for 2011 and 2013.

Describe the revised process for matching HUD-assisted housing units to the AHS national longitudinal sample for 2011 and 2013.

Describe the HUD-assisted variables in the 2011 and 2013 AHS national longitudinal sample public use files (PUFs) and how to use them. The result of the revised processes is greatly improved AHS estimates of HUD-assisted housing units, which provide data users with more precise estimates of key characteristics of HUD-assisted units and HUD-assisted households.

Section 1 discusses the issues discovered after the initial publication of the 2011 national longitudinal sample PUF in 2012, which included the HUD-assisted oversample. Section 1 also discusses issues discovered after matching 2011 HUD-assisted administrative data with the 2011 national longitudinal sample.

Section 2 describes the revised HUD-assisted oversample process, including frame development, selecting the HUD-assisted oversample units, and weighting adjustments necessary for combining the 2011 and 2013 AHS national longitudinal samples with the HUD-assisted oversample.

Section 3 discusses the process of matching HUD-assisted administrative records to the 2011 and 2013 AHS national longitudinal samples. The goal of this process was to identify remaining HUD-assisted housing units within the respective AHS national longitudinal samples. It is important to note that this process is entirely separate from the process that Section 2 describes.

Section 4 describes the development and proper use of the PUF weighting variable *SUBWGT*, which is used to produce estimates of the characteristics of HUD-assisted units and households.

What Users Need to Know

Users who do not wish to read this entire guide should consider reading Section 4.2 and should be aware of the following first three variables in the 2011 national longitudinal sample PUFs and all the following variables in the 2013 national longitudinal sample PUFs.

HUDADMIN

A variable identifying the HUD program to which the unit belongs.

SUBWGT

A weight variable to be used when conducting analysis in which the universe of interest is restricted to HUD-assisted units. All other analysis should be conducted using the weight variable *WGT90GEO*.

SUBWGTRPXXX

Replicate weight variables to be used when calculating variances for estimates using *SUBWGT*.



SP[1/2]SUBWGT

Weight variables to be used when conducting analysis in which the universe is restricted to the HUD-assisted units and topical module variables.

SP[1/2]SUBWGTRPXXX

Replicate weight variables to be used when calculating variances for estimates using for *SP[1/2]SUBWGT*.

These variables are not available for the 2011 or 2013 AHS metropolitan area longitudinal oversample PUFs because the oversample frame was only for the national longitudinal sample.

1. Historical Overview

The following sections discuss issues discovered regarding the 2011 HUD-assisted oversample and efforts to match 2011 HUD-assisted administrative data with the 2011 AHS national longitudinal sample.

1.1. Pre-2011 Administrative Data Matching

Since 1989, HUD has attempted to identify HUD-assisted renters in the AHS national longitudinal sample by address matching AHS housing units to HUD-assisted housing unit administrative records. The HUD administrative records were derived from the Public and Indian Housing Information Center system for Public Housing and Housing Choice Voucher (HCV) Programs and the Tenant Rental Assistance Certification System for project-based Section 8 programs, hereinafter referred to as privately owned multifamily. The matched AHS-HUD data was used to produce *Characteristics of HUD-Assisted Renters and Their Units* reports for 1989, 1991, 1993, and 2009.¹ HUD and the Census Bureau maintained an internal version of a variable indicating whether the AHS record was identified as HUD-assisted as well as a special HUD-assisted weight, which were used to create the *Characteristics* reports. However, HUD and the Census Bureau did not publish a public version of the new variable or weight.

1.2. The 2011 HUD-Assisted Oversample and Integration Issues

For the 2011 AHS, HUD and the Census Bureau added an oversample of 5,259 HUD-assisted housing units to the AHS national longitudinal sample. The goal of the oversample was to improve the precision of estimates for HUD-assisted housing unit and household characteristics.

After the initial release of 2011 AHS national longitudinal sample PUF data, which included HUD-assisted program type variable *HUDADMIN*, HUD commissioned a report (Eggers and Moumen, 2014) to investigate the process of integrating the HUD-assisted oversample into the 2011 AHS national longitudinal sample, as well as the resulting estimates of HUD-assisted units and households produced using the 2011 AHS national longitudinal sample PUF. The findings of the report, as well as concerns

¹ HUD did not conduct the data match from 1995 through 2005. HUD conducted the data matching in 2007 but decided not to publish a report due to data quality concerns.



AHS users raised, strongly suggested errors in the process for integrating the HUD-assisted oversample into the AHS national longitudinal sample, including issues with survey weights.²

In response to these concerns, HUD and the Census Bureau evaluated the entire development process for the HUD-assisted oversample, including how the sample frame was created, how sample units were drawn from the sample frame, and how the sample units were integrated into the 2011 AHS national longitudinal sample. The evaluation confirmed the findings of the report and the concerns of the AHS users.

1.3. The 2011 HUD-Assisted Administrative Data Matching and Matching Issues

Congruent to the investigation of issues with the HUD-assisted oversample, HUD and the Census Bureau re-evaluated the process of matching 2011 HUD-assisted housing unit administrative records to the 2011 AHS national longitudinal sample. This evaluation revealed issues with the efficacy of both the HUD-assisted housing unit administrative records and the matching process for both the 2011 and the 2013 AHS national longitudinal samples.

1.4. Efforts to Address the Issues

Based on the two evaluations described previously, HUD and the Census Bureau took three actions. First, they revised the process for integrating the HUD-assisted oversample (discussed in Section 2) into the 2011 and 2013 AHS national longitudinal samples. Second, they revised the process for matching HUD-assisted housing units to the 2011 and 2013 AHS national longitudinal samples (discussed in Section 3). Third, they released a revised version of the 2011 AHS national longitudinal sample PUF (version 1.5) and the 2013 national longitudinal sample PUF (version 1.3) that included a corrected *HUDADMIN* variable and a revised PUF weight *SUBWGT* to be used in analysis of HUD-assisted units and households (discussed in Section 4).

2. HUD-Assisted Oversample Process

The following sections detail all aspects of the revised process for the HUD-assisted oversample, including frame development, selection of the HUD-assisted oversample units, and weighting adjustments necessary for integrating the HUD-assisted oversample into the 2011 and 2013 national longitudinal samples. The result was a corrected final weight *WGT90GEO* for the 2011 AHS national longitudinal sample PUF version 1.5 and later.

For readers interested in the how the original (and incorrect) process compares with the revised process, Appendix A details the changes between the two processes.

² Readers should note that HUD does not necessarily support all the findings in Eggers and Moumen (2014). Due to time and resource limitations, the Eggers and Moumen (2014) report could not be completed and should be considered a draft report.



2.1. HUD-Assisted Oversample Frame Development

The Census Bureau selected the HUD-assisted oversample from a listing of housing unit addresses known to have participated in HUD programs in 2009. This list of all relevant HUD-assisted housing units is referred to as the HUD-assisted oversample frame. By design, the HUD-assisted oversample frame included only three HUD programs: public housing, project-based Section 8,³ and HOME Investment Partnerships. Housing units participating in the Housing Choice Voucher (HCV) program were not included in the frame.⁴

Like most survey frames, the HUD-assisted oversample frame had several limitations. First, and most importantly, the source for the frame—the 2009 administrative records from HUD—contained many unusable addresses that had to be removed. Second, the source for the frame was a snapshot of HUD administrative records in 2009, so it may have lacked new housing units entering HUD programs between the specific processing date for the administrative files in 2009 and the selection of the HUD-assisted oversample; it may have also contained HUD-assisted housing units that were demolished sometime between 2009 and 2011.

Exhibit 2.1 shows the total number of housing unit records obtained from HUD (2.8 million). Of this total, nearly 1.6 million housing unit records contained unusable addresses and were therefore not eligible for inclusion in the oversample.

Exhibit 2.1. Summary of the HUD-Assisted Oversample Frame

HUD Program	Total Housing Units in Raw 2009 HUD Files	Unusable Addresses That Were Deleted	HUD-Assisted Oversample Frame
Public housing	1,198,846	708,042	490,804
Private project-based	1,544,484	876,255	668,229
HOME Investment Partnerships Program	58,317	10,142	48,175
Total	2,801,647	1,594,439	1,207,208

2.2. HUD-Assisted Oversample Selection

The HUD-assisted oversample frame development process yielded 1,207,208 HUD-assisted housing units initially eligible for sample selection. However, some of these units had to be removed from eligibility due to being located in counties that were not part of the 2011 or 2013 national longitudinal sample frames. The national longitudinal sample was originally drawn in 1985 and was a geographically stratified sample. This sampling process meant that certain counties were selected to be representative of all counties in the United States. These counties are referred to as primary sampling units (PSUs). Some

³ HUD programs in this category include Rent Supplement, Section 221(d)(3) Below Market Interest Rate (BMIR), Section 236, Section 202 Supportive Housing for the Elderly, Project-Based Section 8, Moderate Rehabilitation, and other, smaller programs.

⁴ HUD made this decision for two reasons. First, housing units that a HCV household occupied in 2009 may not have a HCV household in later years. Second, due to the size of the program, HCV households and housing units were believed to be well represented from a statistical perspective in the AHS national longitudinal sample.



HUD-assisted housing units in the initial HUD-assisted oversample frame were not eligible for the oversample, because they were in counties outside of AHS PSUs.

Exhibit 2.2 shows the total number of HUD-assisted housing units in the initial oversample, as well as the number of housing units eligible for the final frame (e.g., those units within AHS *PSUs*).

Exhibit 2.2. Initial and Final HUD-Assisted Oversample Frame

HUD Program	Initial HUD-Assisted Oversample Frame	Final HUD-Assisted Oversample Frame
Public housing	490,804	279,657
Private project-based	668,229	485,023
HOME Investment Partnerships Program	48,175	31,898
Total	1,207,208	796,578

The final HUD-assisted oversample of 5,259 housing units was selected using systematic random sampling from an ordered list for which the oversample frame was sorted by state and county.

2.3. HUD-Assisted Oversample Integration and Weighting

Because the 2011 and 2013 national longitudinal samples are representative of all housing units in the United States, they were also representative of all HUD-assisted housing units in the United States, including the housing units in the HUD-assisted oversample frame. This feature of the national longitudinal samples means that the HUD-assisted oversample could not simply be *added* to the 2011 or 2013 national longitudinal samples, which would have resulted in double-counting housing units. Rather, the HUD-assisted oversample had to be *integrated* into both the 2011 and 2013 national longitudinal samples, with consideration of the designs of both frames. To account for the two independent samples representing the same set of housing units, a weighting adjustment was applied. This adjustment averaged the weights of both samples, enabling them to be used together to make one national estimate. The integration and weight process proceeded in four steps.

Step 1: Calculate the Base Weight for HUD-Assisted Oversample Housing Units

The base weight for the HUD-assisted oversample housing units was simply the inverse probability of selection of each oversample housing unit from the HUD-assisted oversample frame. From the oversample frame of 796,578 housing units, 5,259 HUD-assisted oversample housing units were selected. Because all HUD-assisted housing units on the frame had an equal probability of selection ($5,259 / 796,578 = 0.0066$), the base weight was 151.5.

Step 2: Match the HUD-Assisted Oversample Frame to the National Longitudinal Sample

This step involved matching the entire HUD-assisted oversample frame (1,207,208 housing units) to the 2011 and 2013 national longitudinal sample. The matching was conducted by address. For the 2011 national longitudinal sample, the result was 522 cases that overlapped with the HUD-assisted oversample frame.



Step 3: Calculate the Weight Adjustment Factor

The weight adjustment was necessary for integrating the two samples so they could be used together to make one national estimate. Without the weight adjustment, HUD-assisted housing units would have been double counted, resulting in approximately 1.2 million more HUD-assisted housing units than actually exist.

The weight adjustment simply averaged the size of both the HUD-assisted oversample (5,259) and the housing units within the 2011 and 2013 national longitudinal samples that overlapped with the HUD-assisted oversample frame (522). The weight adjustment for the oversample was approximately $(5,259 / (5,259 + 522) = .901)$. The weight adjustment for the HUD-assisted units within the national longitudinal sample was approximately $(522 / (5,259 + 522) = .09)$.

Step 4: Calculate the Revised Base Weight for the HUD-Assisted Oversample and the HUD-Assisted Housing Units Within the National Longitudinal Sample

The revised base weight was the original base weight multiplied by the weight adjustment factor. For oversample cases, the revised base weight was $(151.47 * .901 = 136.5)$. For housing units within the 2011 and 2013 national longitudinal samples that overlapped with the HUD-assisted oversample frame, a similar calculation was made based on the 2011 and 2013 national longitudinal sample original base weights. Generally speaking, 2011 and 2013 sample base weights ranged from 1,000 to 2,000. For instance, if the national longitudinal sample housing unit base weight was 2,000, the revised based weight was $(2,000 * .09 = 180)$.

2.4. Results and Discussion

The best test of whether the process described in Section 2.3 was successful is to compare the weighted total AHS HUD-assisted housing units with the total housing units from the oversample frame. All else being equal, the weighted total should be equal to the total housing units from the oversample frame.

To make this comparison, the revised base weights were first multiplied by a PSU sampling factor to account for the fact that the 2011 and 2013 national longitudinal samples are geographically stratified. This adjustment facilitates the comparison of the 2011 and 2013 national longitudinal sample estimates of HUD-assisted housing units to the full HUD-assisted oversample frame, which included 1.2 million HUD-assisted units. The third columns of Exhibits 2.3 (for 2011) and 2.4 (for 2013) present the results. Then, the final 2011 and 2013 AHS national longitudinal sample weights were calculated by completing the weighting adjustments, including the nonresponse adjustments and ratio adjustments to other known totals. The final 2011 and 2013 national longitudinal sample estimates with the final weight are in the fourth columns of Exhibits 2.3 and 2.4.

Exhibit 2.3. Comparing the Oversample Frame With 2011 National Longitudinal Sample Estimates

HUD Program	Initial HUD-Assisted Oversample Frame	2011 AHS National Longitudinal Sample with PSU Adjustment	2011 AHS National Longitudinal Sample with All Adjustments
Public housing	490,804	405,190	433,086
Private project-based	668,229	615,162	638,792



HOME Investment Partnerships Program	48,175	35,613	25,124
Total	1,207,208	1,055,964	1,097,001

Exhibit 2.4. Comparing the Oversample Frame With 2013 AHS National longitudinal sample Estimates

HUD Program	Initial HUD-Assisted Oversample Frame	2013 AHS National Longitudinal Sample with PSU Adjustment	2013 AHS National Longitudinal Sample with All Adjustments
Public housing	490,804	409,149	470,862
Private project-based	668,229	614,714	660,955
HOME Investment Partnership Program	48,175	35,480	28,488
Total	1,207,208	1,059,342	1,160,305

The results in Exhibits 2.3 and 2.4 show close agreement between the final 2011 and 2013 AHS national longitudinal sample estimates of HUD-assisted units and the total HUD-assisted units in the oversample frame. The slight difference is likely due to two issues. First, and perhaps most importantly, the process for matching the HUD-assisted oversample frame addresses to the AHS national longitudinal sample may not work perfectly. Recall that part of the cleaning process of the HUD-assisted frame included removing about 57 percent of all HUD addresses, because they were unusable (see Exhibit 2.1). Approximately 1.2 million HUD-assisted housing unit addresses remained that were deemed usable. These addresses were then matched to the AHS national longitudinal sample addresses. This matching process could have failed to flag 2011 or 2013 AHS national longitudinal sample units that were truly HUD-assisted because a) the housing unit address may be slightly incorrect, or b) the matching process may have failed to find a match that truly existed.

The second reason for slight differences between the 2011 and 2013 estimates of HUD-assisted housing units and the HUD-assisted oversample frame is sampling variability (introduced in the first stage of AHS sample selection) or variability due to the selection of HUD-assisted housing units from the oversample frame.

2.5. Conclusion

The results in Exhibits 2.3 and 2.4 show that the revised process for integrating the HUD-assisted oversample into the AHS national longitudinal sample for 2011 and 2013 performs well. The process yields estimates of HUD-assisted housing units that are in close agreement with the total housing units from the HUD-assisted oversample frame. Given these results, HUD and the Census Bureau feel confident in publishing a revised 2011 national PUF.

3. HUD-Assisted Administrative Data Matching Process

This section discusses the process of matching HUD-assisted administrative records to the 2011 and 2013 AHS national longitudinal samples. The goal of matching HUD-assisted administrative records to the 2011 and 2013 AHS national longitudinal samples was to identify remaining HUD-assisted housing units within the AHS national longitudinal sample. It is important to note this process is entirely separate from the process described in Section 2.



3.1. Background

The process of matching the HUD-assisted renters oversample frame to the 2011 and 2013 AHS national longitudinal samples (described in Section 2.3) resulted in the identification of 522 HUD-assisted housing units within the samples. However, this total did not represent all the HUD-assisted units that were expected to exist within either the 2011 or 2013 AHS national longitudinal samples for at least two reasons. First, as discussed in Section 2.1, the HUD-assisted oversample frame, by design, did not include any housing units occupied by households receiving HCVs. Second, the oversample frame was developed based on a 2009 snapshot of HUD administrative records, meaning it is possible that new housing units added to the AHS national longitudinal sample for 2011 or 2013 could have been HUD-assisted households, but would not have matched to the 2009 snapshot⁵.

The intent of HUD and the Census Bureau was to ensure that all HUD-assisted units within the 2011 and 2013 AHS national longitudinal samples were identified. To accomplish that goal, HUD and the Census Bureau performed a match between a 2011-based snapshot of HUD-assisted administrative records and the 2011 AHS national longitudinal sample (and did the same for the 2013 sample). This matching process, which was an entirely separate process from the matching of the HUD-assisted oversample frame to the 2011 and 2013 AHS national longitudinal samples, yielded additional housing units being identified as HUD-assisted units. However, subsequent analysis by HUD and the Census Bureau strongly suggested that the 2011 and 2013 HUD-assisted administrative record matching process failed to identify all the possible HUD-assisted units within the 2011 and 2013 AHS national longitudinal samples.

3.2. Addressing the Issue

HUD and the Census Bureau performed a thorough review of both the 2011 and 2013 HUD-assisted administrative records and the matching process, resulting in three major findings. First, and most importantly, many unit-level addresses HUD initially provided the Census Bureau for 2011 and 2013 were not useable, where usability is defined as an address that is both deliverable by the U.S. Postal Service and unique within the list of addresses. Second, not all 2011 and 2013 AHS national longitudinal sample unit addresses contained sufficient information to ensure address matching would occur correctly. Third, the matching process was constrained to using only physical address information.

HUD and the Census Bureau addressed each of the three issues. First, HUD took steps to substantially improve the quality of address information supplied for the Public Housing program (Bucholtz and Pamnani (2016)) HUD and the Census Bureau developed a technique to improve the unit-level address information for units in the project-based Section 8 program.⁶

Second, the Census Bureau used internal location information to refine certain addresses in the 2011 and 2013 AHS national longitudinal samples helping to ensure that all AHS addresses were at least eligible to be matched to HUD-assisted administrative records.

⁵ New housing units from new construction frame are added to the AHS national longitudinal sample at the beginning of each survey.

⁶ The unit-level address information for HCV units was usable for most records that HUD supplied.



Third, the Census Bureau implemented a revised matching procedure that took advantage of the fact that both the 2011 and 2013 AHS national longitudinal samples and HUD-assisted administrative records contained the names of persons. Matching between the 2011 and 2013 AHS national longitudinal samples and HUD-assisted administrative records used either physical address information or person-level information.

3.3 Using the Administrative Data Match to Verify the Status of the HUD-Assisted Oversample

The HUD-assisted oversample was drawn from a listing of housing unit addresses known to have participated in HUD programs in 2009 (Section 2.1). It was possible, and perhaps likely, that some HUD-assisted oversample units may not have existed in 2011 (or 2013) due to demolition or may have transitioned into the private housing stock.

To account for this possibility, the process described in this section was applied to all HUD-assisted oversample units in the 2011 and 2013 AHS national longitudinal samples. All HUD-assisted oversample units were matched against the 2011 (or 2013) HUD-assisted administrative records to ensure they were still HUD-assisted units. If the HUD-assisted oversample units did not match to the 2011 (or 2013) HUD-assisted administrative records, they were deemed no longer HUD-assisted for the 2011 (or 2013) AHS national longitudinal samples.

3.4 The Results

The final outputs of the processes described in Section 2 and this section are revised values for the *HUDADMIN* variable for the 2011 (version 1.5) and 2013 (version 1.3) AHS national PUFs. The test of whether the processes worked well is to compare the weighted values of *HUDADMIN* with known program totals by using the general AHS national weight *WGT90GEO*.

Exhibit 3.1 shows the total number of 2011 HUD-assisted housing units, by program, using the weight *WGT90GEO*. It also shows known program control totals that reflect the number of HUD-assisted unit, by program, that were known to be occupied on July 1, 2011.

Exhibit 3.1. 2011 AHS HUD-Assisted Totals Using WGT90GEO

HUD Program	Total Housing Units Using WGT90GEO	Known Program Control Total
Public housing	932,989	1,023,178
Private project-based	1,347,953	1,346,314
HCV program	2,239,356	2,097,025
Total	4,520,298	4,466,517

Using *WGT90GEO* produced a weighted HUD-assisted unit estimate of 4,520,298. This estimate is 1.2 percent more than the known total number of HUD-assisted units in 2011 (4,466,517). The HUD-assisted units within the 2011 AHS national longitudinal sample represent about 101.2 percent of the known HUD-assisted units.

Exhibit 3.2 shows the total number of 2013 HUD-assisted housing units, by program, using the regular AHS weight *WGT90GEO*. It also shows known program control totals that reflect the number of HUD-assisted units, by program, that were known to be occupied on July 1, 2013.



Exhibit 3.2. 2013 AHS HUD-Assisted Totals Using WGT90GEO

HUD Program	Total Housing Units Using WGT90GEO	Known Program Control Total
Public housing	1,062,683	1,029,956
Private project-based	1,357,782	1,356,792
HCV program	2,343,624	2,103,243
Total	4,764,090	4,489,994

Using *WGT90GEO* produced a weighted HUD-assisted unit estimate of 4,764,090, which is 6.1 percent more than the known total number of HUD-assisted units in 2011 (4,489,994). The HUD-assisted units within the 2013 AHS national longitudinal sample represent about 106.1 percent of the known HUD-assisted units.

3.5. Removal of *HUDSAMP* from the 2011 and 2013 Public Use Files

HUD and the Census Bureau removed the PUF variable *HUDSAMP* from the 2011 (version 1.5 and later) and 2013 (version 1.3 and later) AHS national longitudinal sample PUFs. The purpose of this decision was to ensure that users do not inadvertently use *HUDSAMP* in their analysis. *HUDSAMP* is present on the 2011 and 2013 AHS national longitudinal sample internal user files (IUF).

4. HUD-Assisted Public Use File Weight *SUBWGT*

This section discusses the creation and use of the HUD-assisted PUF weight called *SUBWGT*, which is weight variable controlled to known HUD-assisted housing unit totals. The *SUBWGT* value for the 2011 AHS national longitudinal sample PUF has been revised from version 1.5. The *SUBWGT* value for the 2013 AHS national longitudinal sample PUF is new for version 1.3.

4.1. Development of *SUBWGT*

Section 2 described how 5,259 HUD-assisted housing units were integrated into the 2011 and 2013 AHS national longitudinal samples, and identified 522 HUD-assisted units already within the 2011 AHS national longitudinal sample. Section 3 described how additional HUD-assisted units were identified within the 2011 (or 2013) AHS national longitudinal samples through matching with the 2011 (or 2013) snapshot of HUD-assisted renters. After accounting for nonresponse, 6,116 HUD-assisted housing units had completed interviews for 2011. For 2013, the total was 6,479. The increase for 2013 can be attributed to the fact that the 2013 AHS national longitudinal sample included the Big 5 metropolitan areas (Chicago, Detroit, New York, Northern New Jersey, and Philadelphia)⁷. Each of these metropolitan areas includes a significant number of HUD-assisted housing units.

The units flagged as HUD-assisted in the 2011 and 2013 AHS national longitudinal sample PUFs can be used to produce housing unit and household characteristics estimates by using the general survey weight *WGT90GEO*. As Section 3.5 showed, estimates produced using the 2011 AHS national longitudinal

⁷ The Big 5 group of metropolitan area longitudinal oversamples do not follow the Office of Management and Budget's 1993 or 2003 metropolitan statistical area boundaries. Users are encouraged to consult Metropolitan Area Oversample Histories: 1973 – 2013 AHS Help Guide.



sample PUF show overcounting of the known total by 1.2 percent. Estimates produced using the 2013 AHS national longitudinal sample PUF show overcounting of the known total by 6.1 percent.

To overcome the undercounting or overcounting issues, HUD and the Census Bureau developed a weighting variable intended to produce housing unit and household characteristic estimates for HUD-assisted households only. This weight variable is called *SUBWGT* and is present on both the 2011 and 2013 AHS national PUFs.

To create *SUBWGT*, HUD and the Census Bureau applied ratio adjustments to the final 2011 and 2013 AHS national longitudinal sample PUFs weight *WGT90GEO*. The ratio adjustments were based on known HUD program control totals of occupied HUD-assisted housing units that HUD provided for both 2011 and 2013. The ratio adjustments were applied separately by the combination of the Census Divisions on the PUFs, two age categories (head of household aged 62 or older or head of household aged less than 62 years), and the HUD program type (public housing, project-based Section 8, or HCV program). Exhibit 4.1 shows the total number of HUD-assisted housing units, by program, using the regular AHS weight *WGT90GEO* and the new weighting variable *SUBWGT*. Exhibit 4.2. contains the 2013 totals.

Exhibit 4.1. Comparison of 2011 AHS National Longitudinal Sample PUF HUD-Assisted Totals Using WGT90GEO and SUBWGT

HUD Program	Total Housing Units Using WGT90GEO	Known Program Control Total	Total Housing Units Using SUBWGT
Public housing	932,989	1,023,178	1,023,178
Private project-based	1,347,953	1,346,314	1,345,953
HCV program	2,239,356	2,097,025	2,097,025
Total	4,520,298	4,466,517	4,466,156

Exhibit 4.2. Comparison of 2013 AHS National Longitudinal Sample PUF HUD-Assisted Totals Using WGT90GEO and SUBWGT

HUD Program	Total Housing Units Using WGT90GEO	Known Program Control Total	Total Housing Units Using SUBWGT
Public housing	1,062,683	1,029,956	1,029,963
Private project-based	1,357,782	1,356,792	1,356,788
HCV program	2,343,624	2,103,243	2,103,243
Total	4,764,090	4,489,991	4,489,995

The advantage of the *SUBWGT* is that it provides improved point estimates of the national HUD-assisted housing unit totals because estimates derived from these weights are consistent with actual totals of HUD-assisted housing units from HUD. A limitation of the *SUBWGT* is that the associated variance estimates cannot be produced.

4.2 Use of HUDADMIN and SUBWGT

HUD and the Census Bureau make the following recommendations concerning the use of *HUDADMIN* and *SUBWGT* for 2011 and 2013:



Recommendation 1

When conducting analysis of the entire housing stock and where “HUD-assisted” is a descriptor of housing units, use the weight variable *WGT90GEO*. For instance, when producing estimates for Exhibit 4.3, the weight variable *WGT90GEO* should be used.

Exhibit 4.3. Example of When to Use *WGT90GEO*

Type of Structure	HUD-Assisted Renters	Nonassisted Renters	Owners
Single-family detached			
Single-family attached			
Multifamily			
Other			

Recommendation 2

When conducting analysis of only HUD-assisted units, use the weight variable *SUBWGT*. For instance, when producing estimates for Exhibit 4.4, the weight variable *SUBWGT* should be used.

Exhibit 4.4. Example of When to Use *SUBWGT*

Type of Structure	Public Housing	Private Project-Based	Vouchers
Single-family detached			
Single-family attached			
Multifamily			
Other			



Appendix A

A.1 Comparison of Original and Revised Oversample Integration Method

The HUD-assisted oversample integration was revised based on three improvements: 1) revised base weights; 2) more accurate matching of the HUD-assisted oversample frame to the 2011 and 2013 AHS national longitudinal samples; and 3) an improved statistical method for combining the 2011 and 2013 AHS national longitudinal samples with the HUD-assisted oversample, resulting in a correct final weight *WGT90GEO*.

First, the original 2011 AHS national longitudinal sample weights assigned to the HUD-assisted oversample used average base weights derived from the 2011 AHS national longitudinal sample. Because the sampling fraction for the HUD-assisted oversample was much larger than the 2011 AHS national longitudinal sampling fraction, those HUD-assisted oversample base weights should have been smaller than 2011 AHS national longitudinal sample base weights.

The revised base weight creation process, as Section 2.3 discussed, fixed this issue. Exhibit A.1 indicates the impact of the revised process. The original release of the 2011 PUF, in which the HUD-assisted oversample base weights were incorrect, yielded about 11.3 million total HUD-assisted housing units. That number is nearly 10 times as large as the HUD-assisted oversample frame. The revised base weight creation process yielded estimates closer to the total HUD-assisted housing units on the HUD-assisted oversample frame.

Exhibit A.1. Comparison of Base Weighted HUD-Assisted Unit Counts for Original and Revised Releases

AHS Year	Frame Total	Base Weight Estimated Total	Sample Size
2011 (original release)	1,207,208	11,296,955	5,259
2011 (revised release)	1,207,208	1,054,123	5,259
2013	1,207,208	1,054,123	5,259

The second key improvement revised the matching of the HUD-assisted oversample frame to the 2011 and 2013 national longitudinal samples. The HUD-assisted oversample frame records were matched with the Master Address File (MAF) and a MAF Identifier (MAFID) from the MAF was assigned to all matches. However, not all the records of the HUD-assisted oversample frame could be matched to the MAF. HUD and the Census Bureau augmented the matching process with additional address information, which improved the identification of HUD-assisted oversample frame housing units within the AHS national longitudinal sample.

Exhibit A.2 shows the impact of the improved matching process. The first row of Exhibit A.2 shows the weighted number of HUD-assisted oversample frame units within the 2011 AHS national longitudinal sample, for which the identification was based on the original address matching. Row two shows the weighted number of HUD-assisted oversample frame units within the 2011 AHS national longitudinal sample, for which the identification was based on improved MAF matching. The estimate from the 2011 and 2013 AHS national longitudinal samples of 1,064,336 is closer to the HUD-assisted oversample frame total of 1,207,208 than the original release of 663,554.

Exhibit A.2. Size of Frame and AHS National Longitudinal HUD-Assisted Sample

Year	Frame Total	Base Weight Estimated Total	Sample Size
------	-------------	-----------------------------	-------------



2011 (original release)	1,207,208	663,554	344
2011 (revised release)	1,207,208	1,064,336	552
2013	1,207,208	1,064,336	552

The third key improvement was to revise the methodology for integrating the HUD-assisted oversample with the 2011 and 2013 AHS national longitudinal samples. Specifically, the improvement was a revision to Step 3 described in Section 2.3.

The original weight adjustment simply averaged the total weighted estimate (using base weights) of both the HUD-assisted oversample and the housing units within the AHS national longitudinal sample that overlapped with the HUD-assisted oversample frame. For instance, if the total weighted estimate, using base weights, of the HUD-assisted oversample was 1.1 million, and the total weighted estimate of the housing units within the AHS national longitudinal sample that overlapped with the HUD-assisted oversample frame was 1.3 million, then the weighted adjustment was—

- For the oversample, $(1.3m / (1.1m + 1.3m) = .542)$.
- For the oversample, $(1.1m / (1.1m + 1.3m) = .458)$.

As Section 2.3 mentioned, the revised weight adjustment simply averaged the size of both the HUD-assisted oversample (5,259) and the housing units within the AHS national longitudinal sample that overlapped with the HUD-assisted oversample frame (522). The weight adjustments were—

- For the oversample, approximately $(5,259 / (5,259 + 522) = .901)$.
- For the HUD-assisted units within AHS national longitudinal sample, approximately $(522 / (5,259 + 522) = .09)$.



References

Bucholtz, Shawn, and Pramod Pamnani. 2016. *Constructing Unit-Level Addresses in the Public and Indian Housing Information Center*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. <https://www.census.gov/programs-surveys/ahs/operations-and-administration/redesign-2015/-constructing-unit-level-addresses-in-the-public-and-indian-hous.html>

Eggers, Frederick J., and Fouad Moumen. 2014. *New AHS PUF Information on HUD-Assisted Rental Housing*. Report prepared for the U.S. Department of Housing and Urban Development. Bethesda, MD: Econometrica.

U.S. Department of Housing and Urban Development
Office of Policy Development and Research
Washington, DC 20410-6000



April 2019