



National Sample Case History File: 1985 to 2013

LAST UPDATED: AUGUST 2019

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





Contents

- 1. Overview 3
- 2. A Note About the “Big 6” Metropolitan Area Oversamples 4
- 3. Identified Patterns 4
- 4. Possible Causes 6
- 5. Data Dictionary 7



1. Overview

The purpose of this document is to describe the American Housing Survey (AHS) *National Sample Case History File*. This file tracks *all* national sample AHS cases that were interviewed between 1985 and 2013 and provide basic information on the interview status. It is important to note the National Sample Case History File does NOT include cases that were part of the independent metropolitan area oversamples that were surveyed between 1985 and 2013.

During AHS data collection, cases are categorized as completed interviews or noninterviews. Noninterviews are classified into three groups: type A, B, and C. Types A and B include respondent refusals or abandoned houses with extensive damage that expose them to the elements. Type A and B cases will be interviewed in the subsequent surveys because they are valid housing units. Type C noninterviews represent units that were completely demolished or permanently removed from the housing stock in some way. Type C cases will not be interviewed in subsequent AHS data collection periods. This document will refer to completed interviews, type As, and type Bs collectively as “eligible housing units” and type Cs as “ineligible housing units”.

The goals of the National Sample Case History File are:

to track *all* records that appeared in the national sample public use files (PUFs) at any point between the 1985 and 2013

explain why the case was removed or reinstated if it does not follow a consistent pattern. The case is considered inconsistent if it is dropped from the PUF in a year after it is introduced or if it changes from being an ineligible housing unit to an eligible housing unit in a later year.

The National Sample Case History File has 125,049 records. This is the total number of unique records that are shown on the national sample PUFs between 1985 and 2013.

The National Sample Case History File has a flat file structure, meaning there is one record for each national sample case that was ever present on a national PUF, which can be uniquely identified by the *CONTROL* variable. This file contains a variable for each year with the *STATUS*, *PWT*, *NOINT*, and *TENURE* codes for that year. If the value is missing, it means the case was not present on a PUF for that year.

The aforementioned variables were filled directly from the PUFs. There are a few notable alterations that were made to these variables.

1. Before 1997, the *STATUS* variable was coded differently than in later years. For this reason, the *STATUS* variables use what was *ISTATUS* on the PUF in order to maintain consistency throughout the file.
2. There are a few records that went out as type Cs the same year they were introduced in 2005 without showing an initial *PWT* value. The *PWT05* variable for these records were adjusted to show their basic weight.
3. Variables for older years were altered to keep them consistent with current formatting practices. This includes changing the variable between character and numeric, applying uniform blanking edits, and adjusting for minor coding differences (for instance between a missing and a B code). The most recent formatting rules for all variables were applied across the board.



4. Prior year type C records were released on a separate file in 2011, so all information concerning these records in those years was taken from that file rather than the PUF.

Additionally, there is a *YRINTRO* code that gives the year the case was first introduced and a *YEARMC* code that gives the year the case was first made an ineligible housing unit, if it was at all.

Finally, the *PATTERN* variable describes what happened to the case over time, and if the case shows an inconsistent pattern across time it describes what happened to cause that. Below is an outline of the different categories for *PATTERN*. For additional information about its coding, refer to the Coding Scheme section of this document.

2. A Note About the “Big 6” Metropolitan Area Oversamples

Although the National Sample Case History File does not include AHS records from independent metropolitan area oversamples, there are three notable exceptions. The three exceptions were:

The AHS independent metropolitan area sample microdata records from the Big 5¹ (Chicago, Detroit, New York City, Northern New Jersey, and Philadelphia) metropolitan areas were integrated into the national longitudinal sample PUF for 2009 and 2013. These cases appear in the National Sample Case History File.

The AHS independent metropolitan area sample microdata records from Los Angeles was integrated into the national longitudinal sample PUF for 2011. These cases appear in the National Sample Case History File.

The AHS independent metropolitan area sample microdata records from the Big 6² areas (the Big 5 plus Los Angeles) was integrated into the national longitudinal sample PUF for 1995, 1999, and 2003. These cases appear in the National Sample Case History File.

3. Identified Patterns

Consistent Pattern

These records show an expected pattern over time. They are present on every PUF after their initial year of introduction. If they become an ineligible housing unit in a certain year, they remain ineligible every year after.

1987 Type C Drop

Returning ineligible housing units were initially dropped off of the file in 1987. All of these records returned as ineligible housing units in 1989 and most stayed that way through to 2011. The

¹ 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.

² The Big 6 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.



exception are 91 records classified as a 1987 Type C Drop that were also dropped from the 2007 file and kept off of 2009 and 2011 due to the 2007 Type C Drop.

1987 Sample Drop

Some eligible housing units were missed in the 1987 interview process but recovered in 1989 and later. The Census Bureau is currently researching the reason behind this.

1997 Sample Drop

Some records were removed from the sample in 1997 only to return in 1999. The Census Bureau is currently researching the reason behind this.

2007 Type C Drop

Some housing units that were made ineligible in 2005 or earlier were removed from the 2007 file and did not appear again in later years.

2007 Sample Drop

These records were reduced from the sample in 2007 for budget reasons. They were all reinstated in 2011 as part of the larger sample.

Big 6 Oversample

These records were part of a supplemental sample that represented the Big 6³ metropolitan area oversamples on the 1995, 1999, and 2003 files. They are a supplemental sample that includes the Chicago, Detroit, Los Angeles, New York, Northern New Jersey, and Philadelphia Metropolitan Statistical Areas. These records acted as national sample records but were added so that estimates for the six aforementioned metro areas could be provided in those years. While the majority of these records are solely used for the Big 6 areas, there does exist some overlap with preexisting rural oversample. Records that were in both samples and were shown as rural and part of a Big 6 area on the PUF will be under this classification rather than that of the rural oversample.

Neighborhood Oversample

For 1985, 1989, and 1993, the AHS sample included a neighborhood oversample. To create a neighborhood oversample, HUD and the Census Bureau first selected existing AHS records. Then, the AHS collected information on up to 10 of the neighbors of the selected AHS records. These up-to-10 additional records for each selected AHS record are not considered part of the longitudinal AHS sample and they only exist in the three aforementioned survey years. There are 20 records in this category that were affected by a sample drop.

Rural Oversample

³ The Big 6 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



These records are part of a special rural oversample that took place in 1987 and 1991 and only appear on PUFs for those years.

Other

There are a number of other reasons for a pattern to not be consistent. The Census Bureau is currently researching what is happening to many of these records, but some of what is happening with the records in this category happen on a case by case basis.

4. Possible Causes

Below are a few possible causes that could fit with many of these records:

Multiple Patterns

Some of these records are affected by multiple of the above patterns. The majority of records fit into one pattern by itself, but for those that were affected by multiple sample drops, type C drops, or any additional reasons outlined below, it is difficult to classify any one reason for the case being inconsistent.

Additional and Extra Units from Before 1997

The AHS has not always treated additional and extra units found during listing or interviewing consistently. In more recent years, they are generally interviewed and added to the sample. Before the electronic data collection instrument, in the paper questionnaire there was a question where the lister could classify additional units as "converted to more", "merged to fewer", "other, Census duplicate", and "no longer living quarters". Any of these choices would set the case to a value of 19 for *NOINT*. If the case resulted in a new or altered unit that was interviewed, it would be added to the file under a new control number. A number of these records were eventually made into a *NOINT* of 37, or other type C. The Census Bureau is currently researching the details surrounding these records.

Changes from Ineligible to Eligible Housing Units

A number of these changes happen on a case by case basis. Though the Census Bureau does not generally reclassify eligibility status once it is put onto a PUF, analysts in the past have had the ability to alter records that were obviously misclassified. The Census Bureau is continually researching individual records to determine how and why they changed in this way.

Data Falsification

There currently exists a number of mechanisms to watch for potential data falsification issues during the interviewing process. These include programs that track how long a case took as well as a number of different indicators that can be sorted by FR. Not all of these tools existed in earlier survey years. The Census Bureau is still researching how records that were potentially falsified were treated and looking into the possibility that it could have had an effect on the continuity of the file.

Methodology Changes

The 1985-2011 survey period covered a number of radical changes in survey methodology. This includes a transition from a strictly paper questionnaire to one that was electronic only. The Census



Bureau is currently researching records that could have been affected by this transition, particularly between 1995 and 1997.

5. Data Dictionary

CONTROL

Control Number

Description

This is the scrambled control number from the master file. It is used to preserve confidentiality and to enable longitudinal matches to earlier files.

Type

Character

PATTERN

Description of what happened to the case over time

Values:

- 01—Consistent and Complete
- 02—1987 Type C Drop
- 03—1987 Sample Drop
- 04—1997 Sample Drop
- 05—2007 Type C Drop
- 06—2007 Sample Drop
- 07—Big 6 Oversample
- 08—Neighborhood Oversample
- 09—Rural Oversample
- 10—Other

Description

For additional documentation on each of these patterns, refer to the “Identified Patterns” section of this document.

Type

Character

NOINTXX

Reason for noninterview

Values

Missing - Case was not on the file in this year



- B—Not applicable, Interviewed case
- 1—Type A - No on home
- 2—Type A – Temporarily absent
- 3—Type A – Refused
- 4—Type A – Unable to locate
- 5—Type A – Language Problem
- 6—Type A – Other, occupied
- 10—Type B – Permit granted, construction not started
- 11—Type B – Under construction, not ready
- 12—Type B – Permanent or temporary business or commercial storage
- 13—Type B – Unoccupied site for mobile home or tent
- 14—Type B – OTHER unit or converted to institutional unit
- 15—Type B – Occupancy prohibited
- 16—Type B – Interior exposed to the elements
- 17—Type B – Not classifies above, specify
- 19—Type B – Split or merged since sample selection, unit is included on file under a different control number (only applicable before 1997)
- 30—Type C – Demolished or disaster loss
- 31—Type C – House or mobile home moved
- 32—Type C – Unit eliminated in structural conversion
- 33—Type C – Merged not in current sample
- 36—Type C – Permit abandoned
- 37—Type C – Not classified above
- 38—Type C – Unit eliminated in sub-sampling
- 39—Type C – Unit deleted in prelisting sub-sampling
- 40—Type C – Unit already had a chance of selection
- 41—Type C – Sample reduction for the current survey year
- 42—Type C – Big Six metro supplement

Description

This is the PUF value of *NOINT*. The last two digits of the variable name correspond to the last two digits of the year to which it refers. For example, *NOINT85* represents the *NOINT* value shown on the 1985 PUF.



Type

Numeric

Notes

Code 39 was only used in 2005. Code 40 was used starting in 2005. Code 41 was introduced in 2009 in Seattle. Code 42 was introduced in 2011.

PWTXX

Pure weight—inverse of probability of selection

Values

Missing—Case was not on the file in this year

B—Case was not in sample this year, returning type C

00000:99999 – 0-99,999

Description

PWT is the product of base weight, weighting control factors, and sample adjustment factors.

This is the PUF value of *PWT*. The last two digits of the variable name correspond to the last two digits of the year to which it refers. For example, *PWT85* represents the *PWT* value shown on the 1985 PUF.

Type

Numeric

STATUSXX

Interview status

Values

Missing—Case was not on the file in this year

1—Occupied interview

2—URE (Usual Residence Elsewhere) interview

3—Vacant interview

4—Noninterview

Description

This is the PUF value of *STATUS*. The last two digits of the variable name correspond to the last two digits of the year to which it refers. For example, *STATUS85* represents the *STATUS* value shown on the 1985 PUF.

Type

Character

TENUREXX



Owner/renter status of unit

Values

Missing—Case was not on the file in this year

B—Not applicable

1—Owned or being bought by someone in your household

2—Rented

3—Occupied without a payment of rent

Description

This is the PUF value of *TENURE*. The last two digits of the variable name correspond to the last two digits of the year to which it refers. For example, *TENURE85* represents the *TENURE* value shown on the 1985 PUF.

Type

Character

YEARMC

Year made Type C

Values

Missing—Case was not made a type C

1985—2011 – 1985 – 2011

Description

This refers to the year the case was made an ineligible housing unit (a type C). If the case ever changes back to an eligible housing unit and later into an ineligible housing unit again, the first year it was shown on the PUF as ineligible is used.

Type

Character

YRINTRO

Year introduced

Values

1985—2011 – 1985 - 2011

Description

This refers to the year the case was first introduced to the sample. If the case moved in and out of sample, the first year the case is on a PUF is used.

Type

Character

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



August 2019