



AMERICAN HOUSING  
SURVEY

# Public Use File Geography: 1985 to 2013

LAST UPDATED: APRIL 2019

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



PD&R



## Contents

1. Overview .....	1
2. Definitions .....	1
3. Background .....	2
4. The National Longitudinal Sample, 1985 – 2013 .....	2
5. Metropolitan Area Samples, 1985 – 2013.....	5
Appendix A.....	8
Geography in the National and Metropolitan Area Summary Statistics Tables .....	8
Appendix B .....	8
Geographic Indicator Confidentiality and Disclosure Avoidance .....	8

## List of Exhibits

Exhibit 4.1. National Sample Geographic Vintages of <i>METRO</i> and <i>METRO3</i> , 1985–2009 ....	3
Exhibit 4.2. National Longitudinal Sample Geographic Indicators Subject to Disclosure Avoidance Techniques .....	5
Exhibit 5.1. Geographic Indicators in the Metropolitan Area Samples, 1985–2013.....	6
Exhibit 5.2. Metropolitan Area Sample Geographic Indicators Subject to Disclosure Avoidance Techniques, 1985 – 2013 .....	7



## 1. Overview

The purpose of this document is to provide detailed information about the geographic indicator variables included in the American Housing Survey (AHS) from 1985 – 2013. The AHS is a longitudinal survey of housing units that was begun in 1973.

As with most housing or demographic surveys, the AHS public use file (PUF) microdata includes geographic indicators so users can create custom tabulations based on their areas of interest. Geography in the AHS PUF microdata can be confusing to many users, for at least a few reasons. First, the AHS is composed of a national longitudinal sample and separate metropolitan area longitudinal oversamples (hereinafter referred to as metropolitan area samples) and they contain different geographic indicators. Second, the vintages for some of the geographic indicators have changed over time (e.g., 1983 metropolitan area samples versus 2003 metropolitan area samples). Finally, to protect respondent confidentiality, geographic disclosure avoidance techniques have been applied to most geographic indicators in the AHS PUF microdata.

To provide clarity, this document separately addresses (1) the 1985 – 2013 national longitudinal sample; and the (2) independent metropolitan area samples surveyed during 1985 - 2013.

For each sample, our intent is to answer two key questions:

- What geographic indicators are available?
- What is the priority order for determining which geographic indicators receive disclosure avoidance?
- When and how are disclosure avoidance techniques applied?

This document includes two appendices. The first appendix describes the geographic indicators that appear in the summary tables, which is a separate (and much less complicated) issue from the geographic indicators in the PUF microdata. The second appendix provides a general overview of geographic disclosure avoidance techniques.

## 2. Definitions

### AHS Metropolitan Area

A metropolitan area, as defined for use in the AHS. In most instances, this is the same as the Office of Management and Budget (OMB) Metropolitan Statistical Area or Metropolitan Division.

### Geographic Vintage

The year of origination for a geographic entity. For instance, 1973 OMB metropolitan area definitions or Census Bureau 1980 urban area definitions.

### Geographic Sliver

This occurs when two geographic entities are overlaid, resulting in the creation of entities belonging to neither, either or both geographic entities. For example, the overlay of OMB metropolitan area sample and Census Bureau urban areas creates four slivers: metro/urban, metro/non-urban, nonmetro/urban, and nonmetro/nonurban.

### Geographic Disclosure Avoidance Technique



The method used to protect the confidentiality of the physical location of the respondent household. There are three techniques: pseudocoding, alteration, and suppression.

### Geographic Disclosure Avoidance Priority List

Reflects the priority ordering of geographic indicators when disclosure avoidance techniques are applied.

## 3. Background

Since its establishment in 1973, the AHS has been composed of a national longitudinal sample and metropolitan area samples. The national longitudinal sample and metropolitan area samples generally received the same questionnaire. In 1985, the Department of Housing and Urban Development (HUD) and the Census Bureau drew a new national longitudinal sample, which was surveyed every two years from 1985 to 2013. The metropolitan area samples were drawn at different times and surveyed at different times.

Traditionally, public use file (PUF) microdata for the national longitudinal sample and metropolitan area samples were provided in separate files; one file containing the national longitudinal sample cases and one file containing all the metropolitan area sample records for every metropolitan area sample conducted in that survey year

## 4. The National Longitudinal Sample, 1985 – 2013

The national longitudinal sample, which was surveyed from 1985 through 2013 was supplemented each year with new construction and other samples of cases reflecting characteristics of special interest, including cases in the Big 5<sup>1</sup> or Big 6 group of metropolitan area samples <sup>2</sup>.

### 4.1 What geographic indicators are available?

Between 1985 and 2013, there were five AHS geographic identifiers on each of the national longitudinal sample PUFs: *SMSA*, *CMSA*, *REGION*, *METRO/METRO3*, and *DEGREE*.

The AHS geographic indicators *SMSA* and *CMSA* are used to identify specific metropolitan areas. *SMSA* refers to standard metropolitan statistical area, which is either a 1983 OMB metropolitan statistical area *MSA* or 1983 OMB primary metropolitan statistical area *PMSA*. *CMSA* refers to a 1983 OMB consolidated metropolitan statistical areas, which is a group of 1983 OMB primary metropolitan statistical areas.

---

<sup>1</sup> The Big 5 group of metropolitan area samples 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".

<sup>2</sup> The Big 6 group of metropolitan area samples 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."



**User Note #1**

For 1985 through 2009 and for 2013, the *SMSA* and *CMSA* indicators on the national sample PUF represent the *MSA*, *PMSA*, or *CMSA* as defined by OMB in 1983.

*REGION* is based on census regions, which remained unchanged between 1985 and 2013.

The history of the *METRO* and *METRO3* indicators is more complex. The purpose of *METRO* and *METRO3* was to introduce central city and urban area geography into the PUF. These geographic indicators have historically been important to HUD and other users of the AHS.

Between 1985 and 1995, the national longitudinal sample PUFs included the *METRO*<sup>3</sup> indicator. *METRO* was based on the geographic combination<sup>4</sup> of:

- 1980 Census Bureau urban Areas
- 1980 Census Incorporated Places outside of Census Bureau urban areas with population greater than 2,500 (referred to as “other urban”)
- 1980 Central Cities of the 1983 OMB *MSAs* and *PMSAs*
- 1983 OMB Metropolitan Statistical Areas or Primary Metropolitan Statistical Areas

The combination of these geographic entities produced seven unique geographic categories within the *METRO* indicator (see Exhibit 4.1).

Starting with the 1997 national longitudinal sample PUF, the indicator *METRO* was replaced with *METRO3*. *METRO3* was built along the same principles as *METRO*, but due to disclosure issues, the *METRO3* indicator has similar, but fewer categories than *METRO*, making it more difficult to identify outlying urban areas within metropolitan areas (see Exhibit 4.1).

**User Note #2**

For the national longitudinal sample from 1985 through 2009, the *METRO3* geographic indicator is based on 1983 *OMB MSA*, *PMSA*, or *CMSA*, and well as 1980 Census Bureau urban areas and 1980 Incorporated Places.

**Exhibit 4.1. National Sample Geographic Vintages of *METRO* and *METRO3*, 1985–2009**

	Geography Vintage	METRO codes (1985 – 1995)	METRO3 codes (1997 – 2013)
Central cities of metropolitan areas	1983 OMB	1	1
Inside metropolitan area, but not in central city – urbanized area	1983 OMB Metropolitan Areas and 1980 Census Bureau Urban Areas	2	2
Inside metropolitan area, but not in central city – other urban	1983 OMB Metropolitan Areas, 1980 Census Bureau Urban Areas, and 1980 Census Incorporated Places outside of Census Urbanized Areas with population greater than 2,500	3	

<sup>3</sup> This is not to be confused with the *METRO* indicator on the metropolitan areas PUFs, which is discussed in section 5 and 6.

<sup>4</sup> In geography terminology, this is referred to as a union.



	Geography Vintage	METRO codes (1985 – 1995)	METRO3 codes (1997 – 2013)
Inside metropolitan area, but not in central city - rural	1983 OMB Metropolitan Areas and 1980 Census Bureau Urban Areas	4	3
Outside metropolitan areas, urbanized		5	4
Outside metropolitan area, other urban		6	
Outside metropolitan area, rural		7	5

The last geographic indicator on the national longitudinal sample PUF is *DEGREE*. *DEGREE* represents the average heating and cooling days based on the location of the housing unit. While *DEGREE* is not purely a geographic indicator, its underlying structure is inherently spatial. As such, it is treated like a geographic indicator when evaluating geographic indicator confidentiality.

Lastly, there are a handful of variables in the national PUF that are socioeconomic indicators, but have values that are based on a geographic indicator. These are the variables representing fair market rents, income limits, and area median income. For *SMSAs* identified by the *SMSA* or *CMSA* geographic indicators, fair market rents and income limits are likely unaltered from their actual values. However, for areas outside of the *SMSAs* identified by the *SMSA* or *CMSA* geographic indicators, the values may be altered to protect confidentiality.

### What is the priority order for determining which geographic indicators receive disclosure avoidance?

To maintain respondent confidentiality, disclosure avoidance techniques were applied to all geographic indicators except *REGION* on the national longitudinal sample PUF between 1985 and 2013.

There are two geographic disclosure avoidance priority lists for the national longitudinal sample, depending on which parts of the national longitudinal sample are being used. This is because the AHS national longitudinal sample was integrated with supplemental sample from the Big 5 (2009 and 2013) or Big 6 (1995, 1999, and 2003) metropolitan areas in certain years. It was the desire of AHS managers to preserve the specific metropolitan area name for the Big 5 and Big 6 cases. The geographic disclosure priority lists are:

National longitudinal sample cases outside of the Big 5 or Big 6 metropolitan areas

- METRO/METRO3
- REGION
- SMSA/CMSA
- DEGREE

National longitudinal sample cases inside of the Big 5 or Big 6 metropolitan areas

- SMSA/CMSA
- REGION
- METRO/METRO3
- DEGREE





## When and how are disclosure avoidance techniques applied?

To maintain respondent confidentiality, disclosure avoidance techniques were applied to the geographic indicators on the national longitudinal sample PUFs. Exhibit 4.2 lists the geographic indicators and the techniques applied to them.

**Exhibit 4.2. National Longitudinal Sample Geographic Indicators Subject to Disclosure Avoidance Techniques**

Geographic Indicator	When: Geographic Confidentiality Issue and Priority	How: Disclosure Avoidance Technique
SMSA	SMSA identifies unique OMB PMSAs or MSAs as they existed in 1983. Many OMB PMSAs and MSAs did not meet the threshold of 100,000 persons.	Suppression: All sample cases in SMSAs where the population was less than 100,000 or outside of SMSAs (nonmetro) were given a value of 9999.
SMSA/CMSA	For SMSAs that are not part of the Big 5 or Big 6, maintaining the true METRO and METRO3 values took precedence over maintaining the true SMSA/PMSA values.	Suppression: Cases in SMSAs where the rural population was less than 100,000 were given a value of 9999. Suppression: Cases in SMSAs where the non-Central City population was less than 100,000 were given a value of 9999.
SMSA	Some cases in Chicago, New York, and Northern New Jersey were added as supplemental sample. Displaying the true SMSA value, would have led to a disclosure violation.	Pseudocode: Some cases in SMSAs in the Chicago, New York, and Northern New Jersey areas were pseudocoded to reflect their location within the general metropolitan area, but not within a specific PMSA. These have SMSA values of 9991, 9992, or 9993.
METRO	For the Big 5 or Big 6 SMSAs, maintaining the true SMSA/PMSA value took precedence over the true METRO value. In some instances, the urbanized area, "other urban" or rural area of these SMSAs contained less than 100,000 persons. As such, disclosing the name SMSA and METRO = 2, 3, or 4 would be a confidentiality violation.	Alteration: For these SMSAs, cases that where METRO = 3 or 4 have been altered to METRO = 2. In some SMSAs, all cases were coded to METRO = 1 or METRO = 2.
METRO3	For the Big 5 or Big 6 SMSAs, maintaining the true SMSA/PMSA value took precedence over the true METRO value. In some instances, the rural area contains less than 100,000 persons. As such, disclosing the name SMSA and METRO3 = 3 would be a confidentiality violation.	Alteration: For these SMSAs, cases that are truly rural (METRO3 = 3) have been altered to (METRO3 = 2).
DEGREE	In some cases, disclosing the true value of DEGREE, when combined with other geographic indicators, would produce a confidentiality violation.	Alteration: Some cases had their DEGREE value altered by replacing the true value (1 – 6) with a value that is as close to the true value as possible without violating confidentiality restrictions.
REGION	None	None

### User Note #3

Using METRO3 = 3 or 5 to identify cases that are rural will produce fairly accurate national results. However, there will be a minor under-count because truly rural cases in the Big 6 metropolitan areas are altered to be urban. Also, a very small undercount will occur because a small number of cases are in a central city, but are rural. These cases are always identified as in a central city. A similar alteration was applied to the METRO variable in 1995.

## 5. Metropolitan Area Samples, 1985 – 2013

In contrast to the national longitudinal sample, the independent metropolitan area samples in the AHS between 1985 and 2013 were drawn at different times and surveyed at different times. Some metropolitan samples were first drawn in the 1974, while others were drawn as late as 2013. Also, in contrast to the national longitudinal sample, the AHS metropolitan area definition for a particular survey year was based



on the most recent OMB MSA definition *at the time of the survey*<sup>5</sup>. Other geographic indicators, such as central city, were sometimes, but not always, based on the most recently available geographic vintage *at the time the sample was originally selected* (see Exhibit 4.3).

To make matters more complicated, the metropolitan area samples also include a geographic indicator called *METRO*, but this indicator had a set of codes with *different* meanings than the *METRO* indicator in the national longitudinal sample that existed between 1985 and 1995.

## What geographic indicators are available?

Exhibit 5.1 describes the type and vintage of the geographic indicators available on the metropolitan area samples from 1985–2013.

### Exhibit 5.1. Geographic Indicators in the Metropolitan Area Samples, 1985–2013

PUF Geographic Variable	Description and Vintage
SMSA	This variable represents the name of the metropolitan area, but the definition of the metropolitan area is based on the <i>OMB MSA</i> or <i>PMSA</i> definitions <u>at the time the sample was drawn</u> , with some exceptions <sup>6</sup> .
PMSA	Based on the 1990 <i>PMSA</i> code for 15 the metropolitan areas surveyed in 1998. This variable is not present for metropolitan areas surveyed in years other than 1998.
PSUDOTCT	Contains a number representing a pseudo Census tract for the year in which the sample was originally drawn. This is not present in all years. Generally speaking, this variable does not correspond with actual Census tracts and should not be used for geographic analysis.
COUNTY	Based on the Census definition of the county at the time the sample was drawn.
STATE	Based on the Census definition of the state at the time the sample was drawn.
METRO	Up to seven codes, representing the primary central city of the metropolitan area, as well as secondary central cities and other areas. The specific central cities and their respective boundaries were usually based on central cities of the <i>OMB MSA</i> or <i>PMSA</i> at the time the sample was drawn. However, when possible, the <i>METRO</i> codes were updated to reflect current central cities and their respective boundaries.
ZONE	Constructed by HUD and usually based on central cities of the <i>OMB MSA</i> or <i>PMSA</i> at the time the sample was drawn. However, when possible, the <i>ZONE</i> codes were updated to reflect current central cities and their respective boundaries.

## What was the priority order for determining which geographic indicators receive disclosure avoidance?

The geographic disclosure priority lists for metropolitan area sample is:

SMSA/CMSA

METRO/METRO3

ZONE/COUNTY/STATE

<sup>5</sup> In the case of the six New England states (ME, VT, NH, MA, RI, and CT), the OMB NECTA definition was used.

<sup>6</sup> The individual metropolitan publication describes the definition used for the AHS metropolitan area for that survey year.





## When and how were disclosure avoidance techniques applied?

To maintain respondent confidentiality, disclosure avoidance techniques were applied to the geographic indicators on the metropolitan area sample PUFs between 1985 and 2009 and for 2013. Exhibit 5.2 lists the geographic indicators and the techniques applied to them.

**Exhibit 5.2. Metropolitan Area Sample Geographic Indicators Subject to Disclosure Avoidance Techniques, 1985 – 2013**

Geographic Indicator	When: Geographic Confidentiality Issue and Priority	How: Disclosure Avoidance Technique
METRO	In some cases, disclosing the <i>METRO</i> value along with other geographic indicators would be a confidentiality violation.	Alteration or Suppression: Some values of <i>METRO</i> have been altered or suppressed.
ZONE	In some cases, disclosing the <i>ZONE</i> value along with other geographic indicators would be a confidentiality violation.	Alteration or Suppression: Some values of <i>ZONE</i> have been altered or suppressed.
STATE	In some instances, the geographic combination of <i>STATE</i> and <i>ZONE</i> created a geographic sliver.	Suppression: Cases where a <i>ZONE</i> spans multiple states have a suppressed <i>STATE</i> code of 99.
COUNTY	In many metropolitan areas, there are counties with less than 100,000 persons. Disclosing the county name would be a confidentiality violation.	Pseudocode: When a specific county cannot be disclosed, it is combined with other counties to form a pseudocounty. If a <i>COUNTY</i> code is above 840, it is pseudo-coded. The full list of <i>COUNTY</i> pseudocodes and what they represent can be found in the AHS Codebook.



## Appendix A

### Geography in the National and Metropolitan Area Summary Statistics Tables

There are two principle data outputs from the AHS during each survey cycle: summary statistics tables and PUF microdata. The majority of the text in this document is meant to assist users of the PUFs in understanding the geographic indicators available on the files. It is very important to note that the geographic indicators on the national and metropolitan area summary statistics tables vary in their vintages. The geographic indicators on the summary tables have varied in the following ways:

- For national summary tables between 1973 and 1982, the metropolitan area geography is based on the 1973 OMB metropolitan area definitions.
- For national summary tables between 1983 and 1991, the metropolitan area geography, including the central city designation, is based on the 1983 OMB metropolitan area definitions while the urban area geography is based on the 1980 Urban Areas by the Census Bureau.
- For national summary tables between 1993 and 2013, the metropolitan area geography, including the central city designation, is based on the 1993 OMB metropolitan area definitions while the Census Bureau urban area geography is based on the 1990 Urban Areas by the Census Bureau. It is important to note that HUD and the Census Bureau did not update the 2003 and later national summary tables with 2003 metropolitan areas or 2000 urban areas.
- The metropolitan area summary tables are based on either the current *OMB* definition of the metropolitan area at the time of publication, or an AHS specific definition. The definitions are explained in the summary table publication files, as well as the *American Housing Survey Metropolitan History: 1974 – 2013*.

## Appendix B

### Geographic Indicator Confidentiality and Disclosure Avoidance

Due to confidentiality restrictions, the geographic indicators on the AHS national longitudinal sample and metropolitan area sample microdata files cannot expose areas with less than 100,000 people<sup>7,8</sup>. This applies to the AHS geographic indicators in three ways:

- Individual geographic indicators (e.g., urban or metropolitan area)
- Combinations of individual geographic indicators that create geographic “slivers” (e.g., urban areas combined with metropolitan areas)

---

<sup>7</sup> This is commonly referred to as the “100,000 persons rule” and applies Census Bureau demographic surveys collected under the authority of U.S. Code Title 13 (which ensures the confidentiality of all survey respondents).

<sup>8</sup> It should be noted that the AHS typically lacks the necessary sample size to make statistically reliable estimates for small areas that exceed 100,000 persons.



- Combinations of different vintages of the *same* geographic indicator that create geographic slivers (e.g., 1980 central city and 1990 central city; 1983 metropolitan areas and 1993 metropolitan areas).

The national longitudinal sample and the metropolitan area samples have historically included different geographic indicators *and* different vintages of the same geographic indicator.

Because of confidentiality restrictions, when the individual geographic indicators or combination of geographic indicators (by type or vintage) produces geographic slivers, the AHS national and metropolitan area PUFs require geographic disclosure avoidance.

There are two dimensions to geographic disclosure avoidance: priority and technique. The geographic disclosure priority dimension emerges when there are multiple geographic indicators, such as is the case with the AHS. When the combination of geographic indicators produces geographic slivers, a decision needs to be made regarding which indicator(s) to protect the true value and for which indicators to adopt disclosure avoidance techniques.

To illustrate, consider housing units in the AHS national longitudinal sample that are located in the Atlanta metropolitan area. In the AHS national longitudinal sample, the specific name of metropolitan area is disclosed (Atlanta), as well as the urban/rural status. However, the rural portion of Atlanta contains less than 100,000 people, so geographic disclosure avoidance is necessary. If survey managers sought to always maintain the specific metropolitan area name, then it would necessary to adopt disclosure avoidance techniques for the urban/rural indicator. The AHS has different priority orders for the national longitudinal sample and metropolitan area samples.

The second dimension to disclosure avoidance is technique. The AHS adopts three disclosure avoidance techniques: pseudocoding, alteration, and suppression. *Pseudocoding* includes replacing the true value of a geographic indicator with an alternative value with less geographic specificity. An example is combining multiple counties to create a pseudo-county. The *alteration* technique includes replacing the true value of a geographic indicator with a different value, typically the closest value by geographic proximity that can be shown without violating the 100,000-person rule. The *suppression* technique includes masking the true value of the geographic indicator. The most common masking technique is to set the value to "Not disclosed." For the AHS, alteration is favored over the other methods. However, the specific method used varies from case to case.

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



April 2019