

American Housing Survey
Public Use File Geography:
1985 - 2013

Updated: December 2014

Table of Contents

1. Purpose.....	2
2. Definitions	2
3. Background.....	3
4. The National Sample, 1985 – 2009 and 2013	3
What geographic indicators are available?	4
What is the priority order for determining which geographic indicators receive disclosure avoidance?	6
When and how are disclosure avoidance techniques applied?	6
5. Independent Metropolitan Samples, 1985 – 2009 and 2013	8
What geographic indicators are available?	8
What was the priority order for determining which geographic indicators receive disclosure avoidance?	9
When and how were disclosure avoidance techniques applied?.....	9
6. The Combined National and Metropolitan Sample for 2011.....	10
What geographic indicators are available?	11
What was the priority order for determining which geographic indicators receive disclosure avoidance?	12
When and how were disclosure avoidance techniques applied?.....	13
Appendix 1: Geography in the National and Metropolitan Area Summary Statistics Tables .	15
Appendix 2: Geographic Indicator Confidentiality and Disclosure Avoidance	16

1. Purpose

The American Housing Survey (AHS) has been around since 1973. It is a longitudinal survey and the current longitudinal sample has been in place in 1985. 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 sample and separate metropolitan area sample and they contain different geographic indicators. Second, the vintages for some of the geographic indicators have changed over time (e.g. 1983 metropolitan areas versus 2003 metropolitan areas). Third, to protect respondent confidentiality, geographic disclosure avoidance techniques have been applied to most geographic indicators in the AHS PUF microdata. Finally, for 2011, HUD combined the national and metropolitan area samples, thereby creating additional confusion.

The purpose of this document is to provide detailed information about the geographic indicators and geographic disclosure in the AHS samples from 1985 through 2013. To provide clarity, this document separately addresses (1) the national sample; (2) independent metropolitan area samples; and (3) the 2011 combined national/metropolitan area sample.

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 American Housing Survey (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 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 areas and Census urban areas creates four slivers:

metro/urban, metro/non-urban, non-metro/urban, and non-metro/non-urban.

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 sample and metropolitan area samples. The national and metropolitan area samples generally received the same questionnaire. The national sample, which is currently conducted every two years, is designed to be representative of the national housing stock, while each individual metropolitan sample is designed to be representative of a metropolitan area's housing stock. The AHS metropolitan areas usually, but not always, coincide with the OMB definitions of the MSA.

Traditionally, public use file (PUF) microdata for the national sample and metropolitan area samples were provided in separate files; one file containing the national sample cases and one file containing all the metropolitan area sample cases for every metropolitan area sample conducted in that survey year. The principle reason for issuing separate files is that the geography indicators (variables) on each file are slightly different, reflecting HUD's desire for different types of geographic identifiers for each sample.

For 2011, the AHS greatly expanded its metropolitan coverage; conducting metropolitan samples in 29 separate AHS metropolitan areas. In an effort to streamline the AHS, the national and metropolitan samples were combined into one sample for 2011. This provided two advantages over the previous practice of keeping the samples separate. First, all cases in the 29 metropolitan samples were used for creating national estimates. This improved the statistical precision of national estimates. Second, cases in the national sample that happen to have been in one of the 29 selected metropolitan areas were combined with the metropolitan sample cases to improve the statistical precision of metropolitan area estimates.

4. The National Sample, 1985 – 2009 and 2013

The national sample used from 1985 through 2013 has been in place since 1985, but supplemented each year with new construction and other samples of cases reflecting characteristics of special interest, including cases in the "Big 6" or "Big 5" metropolitan areas¹.

¹ The "Big 6" metropolitan areas include New York, Northern New Jersey, Philadelphia, Chicago, Detroit, and Los Angeles. The "Big 5" metropolitan areas include the same list, minus Los Angeles.

What geographic indicators are available?

Between 1985 and 2009 and for 2013, there were six AHS geographic identifiers on the national 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.

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 sample PUFs included the METRO² indicator. METRO was based on the geographic combination³ of:

- 1980 Census Urbanized Areas
- 1980 Census Incorporated Places outside of Census Urbanized 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 Table 1).

Starting with the 1997 national 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 Table 1).

USER NOTE #2: For the national sample from 1985 through 2009, the METRO3 geographic indicator is based on 1983 OMB MSA, PMSA, or CMSA, and well as 1980 Urbanized Areas and 1980 Incorporated Places.

² This is not to be confused with the METRO indicator on the metropolitan areas PUFs, which is discussed in section 5 and 6.

³ In geography terminology, this is referred to as a union.

Table 1. National Sample Geographic Vintages of METRO and METRO3, 1985- 2009

	Geography Vintage	METRO codes (1985 – 1995)	METRO3 codes (1997 – 2009, 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 Urban Areas	2	2
Inside metropolitan area, but not in central city – other urban	1983 OMB Metropolitan Areas, 1980 Census Urban Areas, and 1980 Census Incorporated Places outside of Census Urbanized Areas with population greater than 2,500	3	
Inside metropolitan area, but not in central city - rural	1983 OMB Metropolitan Areas and 1980 Census 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 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 sample PUF between 1985 and 2009 and for 2013.

There are two geographic disclosure avoidance priority lists for the national sample, depending on which parts of the national sample are being used. This is because the AHS national 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 sample cases outside of the Big 5 or Big 6 metropolitan areas

1. METRO/METRO3
2. REGION
3. SMSA/CMSA
4. DEGREE

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

1. SMSA/CMSA
2. REGION
3. METRO/METRO3
4. 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 sample PUFs. Table 2 lists the geographic indicators and the techniques applied to them.

Table 2. National 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.	<u>Suppression:</u> All sample cases in SMSAs where the population was less than 100,000 or outside of SMSAs (non-metro) 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.	<u>Suppression:</u> Cases in SMSAs where the rural population was less than 100,000 were given a value of 9999. <u>Suppression:</u> 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.	<u>Pseudocode:</u> 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.	<u>Alteration:</u> For these SMSAs, cases that where METRO = 3 or 4 have been altered to METRO = 2. In some SMSA's, 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.	<u>Alteration:</u> 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.	<u>Alteration:</u> 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. Independent Metropolitan Samples, 1985 – 2009 and 2013

In contrast to the national sample, the independent metropolitan samples in the AHS were drawn at different times. Some metropolitan samples have been in place since the 1974, while others are more recent. In contrast to the national 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*⁴. 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 Table 3).

To make matters more complicated, the metropolitan 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 sample that existed between 1985 and 1995.

What geographic indicators are available?

Table 3 describes the type and vintage of the geographic indicators available on the independent metropolitan area sample from 1985 – 2009 and 2013.

⁴ In the case of the six New England states (ME, VT, NH, MA, RI, and CT), the OMB NECTA definition was used.

Table 3. Geographic Indicators in the Independent Metropolitan Area Samples, 1985- 2009 and 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 OMB MSA or PMSA definitions <u>at the time the sample was drawn</u> , with some exceptions ⁵ .
PMSA	Based on the 1990 PMSA 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 OMB MSA or PMSA at the time the sample was drawn. However, when possible, the METRO codes were updated to reflect current central cities and their respective boundaries.
ZONE	Constructed by HUD and usually based on central cities of the OMB MSA or PMSA at the time the sample was drawn. However, when possible, the ZONE 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 independent metropolitan area sample is:

1. SMSA/CMSA
2. METRO/METRO3
3. ZONE/COUNTY/STATE

When and how were disclosure avoidance techniques applied?

To maintain respondent confidentiality, disclosure avoidance techniques were applied to the geographic indicators on the metropolitan sample PUFs between 1985 and 2009 and for 2013. Table 4 lists the geographic indicators and the techniques applied to them.

⁵ The individual metropolitan publication describes the definition used for the AHS metropolitan area for that survey year.

Table 4. Independent Metropolitan Sample Geographic Indicators Subject to Disclosure Avoidance Techniques, 1985 – 2009 and 2013

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

6. The Combined National and Metropolitan Sample for 2011

As mentioned in the Background section, the national and metropolitan samples were combined into one sample for 2011. The strategy for the 2011 AHS PUF geographic indicators was to populate the traditional national and metropolitan area geographic indicators for **all** cases within any of the 29 metropolitan area samples. In other words, the strategy was to populate ZONE, STATE, METRO, and COUNTY for the national cases within the 29 selected metropolitan area and CMSA, METRO3, REGION, DIVISION, and DEGREE for the metropolitan area sample cases within the 29 selected metropolitan areas. This strategy was not applied to cases in the national sample that were outside of the 29 selected metropolitan areas.

As one might imagine, the strategy of populating all geographic indicators for all cases in the 29 selected metropolitan areas created many instances of geographic slivers brought on by varying vintages of geography within the two samples.

To better understand the problems caused by this strategy, consider the national sample and the Atlanta AHS metropolitan sample. National sample cases that were in the Atlanta metropolitan area in 1983 have an SMSA value of 1111. The METRO3 variable includes

urban areas and central cities circa 1980 Census.

In contrast, the current Atlanta metropolitan sample was drawn in 1996 and was used in 1996, 2004, and 2011. During the initial sample in 1996, the Atlanta AHS metropolitan area was based on the 1993 OMB MSA, with a few modifications. The variable METRO was based on the 1990 boundaries for central cities of the 1993 OMB MSA definition. The 2004 Atlanta metropolitan sample was also based on the Atlanta AHS metropolitan area from 1996.

In the 2011 combined sample, the 2011 Atlanta AHS metropolitan area was based on the 2003 OMB Atlanta MSA. As such, the SMSA indicator now reflects the metropolitan area circa 2003. However, the other geographic indicators contain a mix of geographic vintages. Recall that the METRO3 indicator (coming from the national sample) was based on metropolitan area, urban areas, and central city boundaries as they existed in 1980. The METRO and ZONE indicators (coming from the metropolitan sample) were based on 1990 boundaries for central cities and the 1993 OMB MSA definition. Updating all the geographic indicators using 2000 urban area, central city, and metropolitan area definitions would have created a large number of geographic slivers that would have violated confidentiality restrictions. In fact, populating the METRO3 value for metropolitan sample cases (using 1980's geography) would have created geographic slivers from the combination of 1980 and 1990 central city boundaries and metropolitan area boundaries.

Rather than choosing not to disclose any geographic indicators, the 2011 AHS METRO, METRO3, and ZONE geographic indicators for cases within the 29 selected metropolitan areas were populated, where feasible, with either 1990 or 2000 vintages of central cities. The urban area portion of METRO3 is generally based on the 1980 urban area geography. See Table 5.

What geographic indicators are available?

Table 5. Geographic Indicators in the 2011 AHS

PUF Geographic Variable	AHS Combined Sample (2011)
SMSA	<p>For cases not in the 29 selected metropolitan areas: Based on the 1983 OMB definition.</p> <p>For cases in the 29 selected metropolitan areas: Based on the 2003 OMB MSA or Metropolitan Division definitions with some exceptions (see the "American Housing Survey Metropolitan History: 1974 – 2013").</p>
CMSA	<p>For cases not in the 2011 selected metropolitan areas: Based on the 1983 OMB definition.</p> <p>For cases in the 29 selected metropolitan areas: Based on the 2003 OMB MSA or Metropolitan Division definitions with some exceptions (see the "American Housing Survey Metropolitan History: 1974 – 2013").</p>
REGION	Based on the four Census region codes.

DIVISION	Based on the nine Census division codes.
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.
METRO3	For national sample cases not in the 29 selected metropolitan areas: Based on the 1983 OMB definition for metropolitan area, the 1980 urban area geography and the 1980 central city geography. For national sample cases in the 29 selected metropolitan areas: Based on the 1983 OMB definition for metropolitan area, the 1980 Census Urban Areas and the 1980 Central Cities. For metropolitan sample cases in the 29 selected metropolitan areas: Based on the current definition of the metropolitan areas, either 1980's, 1990's, or 2000's central city geography and 1980 urban area geography.
METRO	For national sample cases in the 29 selected metropolitan areas 1980 central cities geography. For metropolitan sample cases in the 29 selected metropolitan areas: Based on either 1980's, 1990's, or 2000's central city geography.
ZONE	Constructed by HUD and based either 1990's, or 2000's central city geography.
DEGREE	Constructed to represent the average heating and cooling days based on the county of the sample unit.

USER NOTE #4: AHS users should generally be confident when comparing 2011 estimates with past survey estimates using the METRO3 geographic indicator. However, users should consider using the "national only" sample weight in order to ensure consistency with past estimates.

USER NOTE #5: AHS users are strongly cautioned against using the METRO3 or METRO variables when comparing 2011 AHS estimates among different metropolitan areas. The vintages of the central city indicators used to create METRO3 and METRO are inconsistent across the 2011 metropolitan area samples.

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

For cases outside of the national sample, the geographic disclosure avoidance priority list is the same as for the national sample 1985-2009 and 2013.

For cases in one of the 29 selected metropolitan areas, the geographic disclosure avoidance priority list for the 2011 combined national and metropolitan area sample is:

1. SMSA/CMSA
2. REGION
3. METRO/METRO3
4. ZONE/COUNTY/STATE
5. DEGREE

When and how were disclosure avoidance techniques applied?

As with previous PUFs, to maintain respondent confidentiality, disclosure avoidance techniques were applied to the geographic indicators on the 2011 PUF. Table 6 lists the geographic indicators and the techniques applied to them.

Table 6. AHS 2011 Geographic Indicators Subject to Disclosure Avoidance Techniques

Geographic Indicator	When: Geographic Confidentiality Issue and Priority	How: Disclosure Avoidance Technique
SMSA, STATE, COUNTY, METRO, METRO3, and ZONE	Some cases in the national sample were, prior to 2003, not within a specific metropolitan area. However, for the 2011 combined sample, these cases fell within one of the 29 selected metropolitan areas. This most often happens when a metropolitan area expanded between 1983 and 2003. In these cases, disclosing the specific metropolitan area would have created a disclosure violation because the metropolitan area expansion added or removed counties with less than 100,000 people, thereby creating a sliver.	<u>Suppression:</u> These national sample cases remained as SMSA=9999, although they are part of a 2011 metropolitan area. In addition, STATE (99), COUNTY (999), METRO (9), METRO3 (9), and ZONE (999) were suppressed.
COUNTY	For national and metropolitan sample cases in the 29 selected metropolitan areas, some counties in have less than 100,000 people.	<u>Pseudocode:</u> When a specific county cannot be disclosed, it is combined with other counties to form a pseudocounty. If a COUNTY code is above 840, it is pseudo coded. The full list of COUNTY pseudocodes and what they represent can be found in the AHS Codebook.
STATE	For national and metropolitan sample cases in the 29 selected metropolitan areas, the combination of ZONE and STATE	<u>Suppression:</u> Cases where a ZONE spans multiple states have a suppressed STATE code of 99.

	leads to slivers of less than 100,000 people.	
DIVISION	The combination of certain DEGREE codes with some DIVISION values showed areas with less than 100,000 people.	<u>Pseudocode:</u> Census Divisions 5 and 6 were combined into "56" and 8 and 9 were combined into "89."
DIVISION/REGION	For metropolitan sample cases in the 29 selected metropolitan areas, some of the 29 selected metropolitan areas crossed Census Region or Division boundaries, thereby creating areas within the metropolitan area than had less than 100,000 people.	<u>Alteration:</u> The Census Region or Division code was altered if the case was in a portion of the metropolitan area that has less than 100,000 people.
METRO	For metropolitan sample cases in the 29 selected metropolitan areas, there are some metropolitan sample cases where showing the true METRO would have caused a disclosure issue, for a variety of reasons.	<u>Alteration:</u> For metropolitan sample cases in the 29 selected metropolitan areas: some METRO values have been altered to a value of an adjacent zone.
METRO3	For metropolitan sample cases in the 29 selected metropolitan areas, the rural area contains less than 100,000 persons. As such, disclosing the name (SMSA) and METRO3 = 3 would be a confidentiality violation.	<u>Alteration:</u> For metropolitan sample cases in the 29 selected metropolitan areas: cases that are truly rural (METRO3 = 3) have been altered to (METRO3 = 2).
ZONE	For metropolitan sample cases in the 29 selected metropolitan areas, showing any ZONE code would have caused a disclosure issue, for a variety of reasons. However, these cases were still eligible to display their CMSA or SMSA values.	<u>Pseudocode:</u> Cases where CMSA and SMSA could be shown had ZONE set to 990, which allows them to still be identified to a specific metropolitan area. <u>Alteration:</u> Some ZONE values have been altered to a value of an adjacent zone.
DEGREE	In some national and metropolitan area sample cases, disclosing the true value of DEGREE, when combined with other geographic indicators, would produce a confidentiality violation.	<u>Alteration:</u> 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.

Appendix 1: 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. However, 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 Census's 1980 Urban Areas.
- 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 urban area geography is based on Census's 1990 Urban Areas. It is important to note that HUD and Census did not update the 2003 and later national summary tables with 2003 metros 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 2: Geographic Indicator Confidentiality and Disclosure Avoidance

Due to confidentiality restrictions, the geographic indicators on the AHS national and metropolitan sample microdata files cannot expose areas with less than 100,000 people^{6,7}. 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);
- 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).

As previously mentioned, the national and metropolitan 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 areas 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 sample that are located in the Atlanta metropolitan area. In the AHS national 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 for the urban/rural indicator. The AHS has different priority orders for the national sample and metropolitan 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

⁶ 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).

⁷ 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.

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.