The American Housing Survey 2015 Redesign: Impact and Analysis

Toni Warner, and Aref Dajani
Longitudinal Surveys Branch, Demographic Statistical Methods Division, U.S. Census Bureau[1]
June 17, 2011

Overview

The American Housing Survey (AHS) is a longitudinal survey of housing units. The current sample design includes housing units selected in 1985 and has been supplemented with new housing units over time to account for new construction, survey participation attrition, and oversampling of selected populations[2]. The current survey design serves a wide variety of stakeholders, including those who make use of a single year of cross-sectional data and those who link current housing units across time. The current sample is scheduled for final interviews in 2013. A new sample will be drawn for the 2015 AHS.

Maintaining the AHS longitudinal sample for nearly thirty years has presented many challenges, including attrition of housing units, response burden, changes in geography, and disclosure avoidance and mitigation. Because a new sample will be drawn for the 2015 AHS, an opportunity exists to modify the survey design to address the challenges while continuing to serve a wide variety of stakeholders.

This white paper identifies and rigorously explores four key challenges with the current survey design: attrition rates, geographic disclosure, maintaining a representative sample, and appropriate sample size. After discussing the four key challenges, the white paper presents four options for the 2015 AHS design. Finally, each of the four key challenges is revisited in the context of how a modified survey design would address the challenge.

Background

In 1985, the American Housing Survey (AHS) launched a longitudinal sample design where selected housing units were to remain in a national sample over time. This has evolved into a thirty-year panel, with final interviews scheduled for 2013. This sample has been supplemented by units selected to account for new construction, survey participation attrition, and oversampling of selected populations. While some surveys conducted by the U.S. Census Bureau focus on the householder, the AHS focuses on the housing unit. When a householder in an AHS housing unit moves, the next AHS survey interviews the in-movers to the housing unit. Out-movers from the housing units are not interviewed.

Killion (2011) presents the AHS 1985 Sample Design in detail. The AHS National and Metropolitan Sample (AHS-N and AHS-MS) frames incorporate elements from the following sources:

[1] This draft report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.
[2] This will include subsidized housing for 2015 and may include oversampling certain metropolitan areas.

- Housing units selected from the 1970, 1980, 1990, and 2000 census. The 1980 Decennial Census is the primary source for the AHS-National sample.
- Housing units selected from the 1980, 1990, and 2000 permit areas for new construction
- Housing units missed in the 1980 census
- Other housing units added since the 1980 census

This design serves a wide variety of stakeholders. Those interested in the up-to-date status of the housing market would look at one year of cross-sectional data. Those interested in incremental change in the housing market, currently two years apart, link successive interviews through unique housing unit identifiers, each in a Public Use File (PUF) specific to its year of interview. Those interested in change over a longer period can link files from many years apart, for those housing units that are completed interviews in both the beginning and end of the period of interest. Finally, those looking for spells of activity in the housing market can link all the PUFs and select those households with completed interviews for every interview from the beginning and end of the period of interest.

Pre-Determined Updates to the 2015 Sample Design

Several options for the 2015 AHS sample design are presented in a subsequent section. Irrespective of which option is chosen, a few changes to the 2015 sample design will be made. These include:

- There will no longer be separate national (AHS-N) and metropolitan area (AHS-MS) surveys. A representative national sample will be surveyed during every survey. The national sample will be supplemented with additional oversampling cases in each of 60 metropolitan areas.
- Metropolitan oversample cases in 30 of the 60 metropolitan areas will be surveyed in 2015 and in every other survey (2019, 2023, and so on). Metropolitan oversample cases in the other 30 of the 60 metropolitan areas will be surveyed in 2017 and in every other survey year (2021, 2025, and so on).
- The entire sample will be used to generate a single AHS Public Use File (PUF) and statistical tables both nationally and for requested metropolitan areas.
- All sample units will be selected from a continuously updated Master Address File (MAF).
- The location of Primary Sampling Units (PSU) will not be published, thereby providing much greater protection against disclosure.
- PUFs will be constructed at geographic level of nine census divisions.
- Estimates will be produced at all geographic levels, including states, subject to disclosure and reliability considerations. This will have an impact on sampling at the PSU level.
- The AHS will join all other demographic surveys in sampling at the county level in New England.

Challenge #1: Attrition Rates

The current AHS longitudinal sample has been in place for nearly thirty years. Although the frequency of interviews for any one housing unit is only once every two years, it is not hard to

believe that a household occupying a housing unit may not want to participate in the AHS during each survey cycle. The rate at which housing units leave the sample permanently or temporarily is known as the attrition rate. Measuring the historical attrition rate for the current sample helps determine the degree to which attrition may be a challenge for the future survey design.

Attrition rates can be measured in various ways and attrition measurements can produce different results. For the purposes of assessing AHS attrition rates, three measures of attrition are calculated:

Permanent Attrition: In this measure attrition is defined as units that were selected into the sample in 1985, became a "non-interview" status at some point, and remained so through 2009. Two types of "non-interview" statuses are:

- Type A- refusals and housing units that field representatives are unable to locate.
- Type B- housing units under construction or an empty mobile home pad.

This measure of attrition provides a conservative estimate. In each case, these sample units are not replaced with new sample, but are followed up from time period to time period, a burden in this case not on the respondents but on the field representatives. Though there is interest in knowing how large this group is, no information is gathered from these units after this attrition occurs.

Survey Year-Specific Attrition: In this measure attrition is defined as units that are permanently in sample since 1985 and do not have a completed interview for the specific year of interest. For instance, the 2001 attrition rate would measure housing units that were selected into the sample in 1985 but could not be interviewed in 2001 for a Type A or Type B reason. This measure is relevant to stakeholders measuring change over a specific time period.

Intermittent Attrition: In this measure attrition is defined as units that are permanently in sample since 1985, but have at least one incomplete interview between 1985 and 2009. This attrition would be critical for stakeholders measuring spells of activity or event-history analysis, such as when one considers "underwater" status over time, where the loan due on a house is greater than its market value.

The three measures of attrition rate are presented in Table 1. In an analysis of attrition rates it is useful to consider both the absolute attrition rate from start to end, as well as the rate of change in the attrition rate over time.

The AHS attrition rate analysis indicates that permanent loss to sample (6.2%) has been relatively minimal over the 24 years since the 1985 design. Perhaps more importantly, nearly two-thirds of the attrition occurred after 2005, suggesting that after 20 years, attrition could be expected to increase at a higher rate.

The survey-year specific attrition rates are considerably higher than the permanent attrition rate, both in absolute and rate of change. However, the absolute value exhibited periods of stability

from 1989 – 1993, from 1997 – 2003, and from 2007-2009. In between these periods of stability are large increases.

The intermittent attrition rates are much larger than the permanent and survey year-specific attrition rates. In absolute terms, nearly half of the sample selected in 1985 could not be surveyed in at least one year between 1985 and 2009. The rate of change in attrition was consistent between 1985 and 2009, increasing by nearly 4 percent each additional survey year.

Table 1. AHS-National Attrition Rates, 1985 - 2009

Year	Permanent	Survey Year-Specific	Intermittent
1985	0.3%	5.5%	5.5%
1987	0.4%	8.1%	8.1%
1989	0.5%	9.6%	11.1%
1991	0.6%	10.3%	14.1%
1993	0.7%	10.2%	16.3%
1995	0.9%	13.1%	20.6%
1997	1.0%	15.6%	25.7%
1999	1.3%	15.3%	29.8%
2001	1.5%	16.0%	33.9%
2003	1.8%	15.3%	36.9%
2005	2.4%	17.6%	40.8%
2007	3.4%	20.0%	45.2%
2009	6.2%	19.3%	48.4%

Challenge #2: Geographic Identifiers and their Impact on Disclosure

The AHS sample is of sufficient size to permit sub-national estimates, including estimates for Census Regions, urban and rural areas, OMB-defined metropolitan areas, and in some cases, HUD-defined sub-areas within metropolitan areas[3]. Census includes various geographic identifiers in the AHS-N and AHS-MS PUFs.

The metropolitan areas identified on the AHS-N PUF in the current AHS sample are defined using the 1983 Office of Management and Budget (OMB) definitions and are based on1980 geography. They have not been updated in AHS since 1985. This is one of the biggest issues for AHS-N users when they are using public use data[4]. Other data sources offer estimates for metropolitan areas using updated definitions and geography, such as the OMB's 2003 definition of metropolitan areas. This makes it especially hard to link and compare the AHS to other survey data.

Census data release is made under Title 13 of the U.S. Code, which prohibits wrongful disclosure

[3] These areas are defined by HUD and use a combination of Census Urban Area designations and expert opinion.
[4] The metropolitan areas included in the current AHS-MS PUFs were defined using either the 2003 OMB definitions (based on 1990 geography) or are based on HUD-specific definitions which may not coincide with OMB.

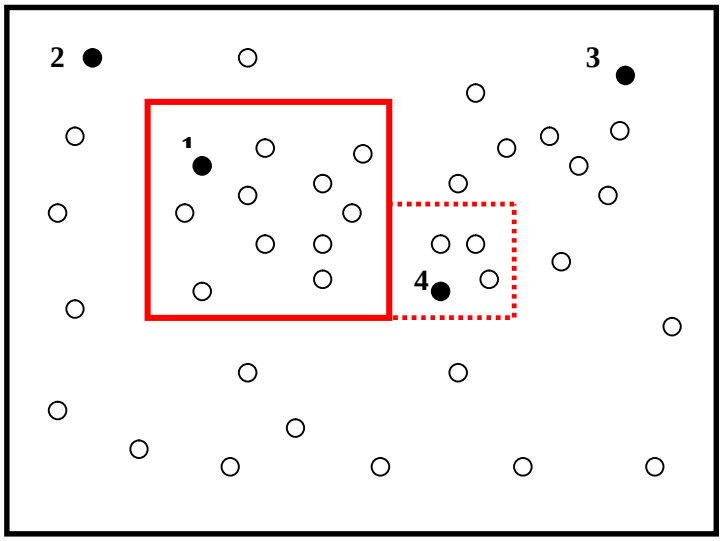
of information on individuals. To make the identification of individuals highly unlikely, the Census does not identify geographic areas with less than 100,000 people in the PUF[5]. Because there are already several geographic identifiers on the AHS-N PUFs, each new geographic identifier added has the potential to violate disclosure standards. This is one of the reasons why AHS-N PUF metropolitan area definitions have not been updated since 1985.

Figure 1 provides a visual representation of a disclosure issue that would arise if a new geographic identifier was added to the PUF. The black box represents a State and each circle represents an AHS sample point. Within the State there is a 1983 metropolitan area represented by the red box. The metropolitan area has 150,000 people, and so AHS sample points within the red solid box included an identifier for the metropolitan area on the 1985 AHS PUF. Points outside the box were identified as not being in a metropolitan area. There is no disclosure violation under this scenario because the metropolitan area includes 150,000 people.

Suppose that the metropolitan area was expanded in 2003 to include the area denoted by the red dashed line. This area includes 25,000 people. If the 2003 AHS were to include a new geographic identifier for the 2003 metropolitan areas (in addition to the identifier for the 1983 metropolitan areas) in the PUF, sample point #4 could be identified geographically as being in an area with less than 100,000 people[6]. This identification on the 2003 AHS PUF would violate Census disclosure rules.

The consequence of the Census disclosure rule is that updated geography is not typically included in the AHS PUFs. In limited circumstances, such as the AHS-MS, some updated geography is provided.

Figure 1. An Illustration of the Impact of Changing Geography on Disclosure



Point	Metro 83	Metro 03
1	YES	YES
2	NO	NO
3	NO	NO
4	NO	YES

[5] Hawala (2001), "Enhancing the '100,000' Rule: On the Variation of the Percent of Uniques in a Microdata Sample and the Geographic Area Size Identified on the File."

[6] The same would be true if the 2003 PUF included only the 2003 metropolitan area. In this case, users could simply link the 1983 and 2003 PUFs longitudanaly, thereby attaching both the 1983 and 2003 metropolitan identifiers to the same file.

The number of disclosure issues created by the inclusion of geographic identifiers on the PUF is directly related to the number of geographic identifiers included. Every new geographic identifier included on the PUF, whether it is a new type of geography or a new vintage of an existing geography, must be intersected with each existing geographic identifier to ensure that small areas cannot be identified.[7]

One possibility for a the 2015 sample design is to reduce the number of geographic identifiers on the current PUF and replace them with other geographic identifiers of interest to the AHS user community, including identifying counties or groups or counties and adding new metro area definitions when they become available.

Challenge #3: Maintaining a Representative Sample

A key to any successful survey design is ensuring the survey sample is representative of the population under study. This is especially true for AHS's longitudinal survey design in which the sample was drawn many years ago and sample replacement is small. If key conditions change in the housing universe and new sample is not added at a rate or type that ensures the total sample is representative of the universe, then conclusions drawn from the sample may be biased.

Census addresses this challenge in two ways. First, as mentioned previously, the current AHS-N and ASH-MS samples are supplemented with new housing units over time to account for new construction. Although the supplemental samples are small, adding newly constructed housing may help ensure that AHS sample reflects the housing universe.

Second, and more important, Census performs several weighting adjustments, including non-interview adjustments, PSU adjustments, new construction adjustments, and demographic adjustments. Each of these adjustments are based in part on measures from other Census surveys, including Decennial Census, Current Population Survey, Housing Vacancy Survey, Survey of Construction, and the Manufactured Homes Survey[8]. The aggregate impact of the weighting adjustments should be a sample that is representative of the current housing universe.

To assess the impact of the weighting adjustments, comparisons can be made between variables that are present in both the AHS and other surveys, assuming that the variables are derived from questions that are the same or substantially similar.

Schwartz (2011) conducted such an analysis that compared estimates from the 2007 AHS and the 2007 ACS for a selected set of variables that were derived from same or similar questions. The 2007 ACS includes a much larger and more current sample than the AHS. Although there are

[7] Census has identified the sub-county variables METRO3 and ZONE as causing the most disclosure related problems. This is due to the multiple types of geographies used to build each of the variables, including new vintages for each of these geographies. For instance, METRO3 is comprised of metropolitan areas, Census urban areas, and Census central cities.

[8] An in-depth explanation of adjustments is provided in Appendix B of the American Housing Survey for the United States: 2009 and is available at: http://www.census.gov/prod/2011pubs/h150-09.pdf

important differences between the two surveys, especially in the manner in which occupancy is measured, they are similar enough to draw general conclusions about how the AHS compares with a larger and more recent survey.

Schwartz took great care to detail why differences in AHS and ACS may exist, including how particular questions are asked and how choice sets are formulated. It was beyond the scope of Schwartz's study to parse reasons why differences in estimates exist into categories, such as differences in questions, differences in answer sets, timing of the survey, or sample bias. It was also beyond the scope of Schwartz's study to draw definitive conclusions about sample bias in the AHS.

An examination of the results, in the context of Schwartz's explanation about each of questions and their corresponding answer sets, may lead the researcher to conclude the there is no systematic sample bias in the AHS. This conclusion may be supported by three pieces of evidence from Schwartz's analysis:

- Estimates from five variables (Units in Structure, Number of Bedrooms, Home Value, Year Householder Moved into Home, Age of Householder) show an inconsistency of statistically significance difference between AHS and ACS among the categories of answers. Furthermore, there is no definitive pattern in the differences. These five variables likely represent the most straight-forward questions and therefore, are the least likely to be impacted by differences in how or when the questions were asked.
- Estimates from the Year Built variable show statistically significant differences in the AHS and ACS estimates for all categories. As Schwartz indicated, there appears to be a slight bias in AHS towards older homes. However, the differences between categories are very small and the difference in median values for this variable for AHS (1973) and ACS (1974) is also small.
- Estimates from the Rooms in Structure variable show statistically significant differences in the AHS and ACS estimates for nearly all categories. As Schwartz indicated, there appears to be a slight bias in ACS towards very small and very large homes. However, the differences are small. Furthermore, the differences may be attributable to minor differences in what AHS and ACS consider to be a "room" for the purposes of the question.

Challenge #4: Achieving an Appropriate Sample Size in an Uncertain Budget Environment

The challenge of achieving an appropriate sample size is more often a matter of budget than of sample design or survey administration. Nevertheless, in a complex survey like the AHS it is important to establish goals for the survey, determine the sample size required to meet those goals, and evaluate survey outcomes to ensure goals are being achieved. Moreover, the sample size is an important component when considering options for re-designing the 2015 AHS. An appropriate sample design is one that is flexible enough to maintain usefulness when budget challenges require sample size changes to be made.

A general goal in the AHS sample design is to produce estimates for 5 percent sub-populations at the national level that have high reliability and precision. HUD has defined high reliability and precision to be percent coefficient of variation (%CV) of five percent.

There are many variables within the AHS that produce sub-populations that are approximately 5 percent. Calculating the %CV for each of these variables is burdensome. Therefore, four variables that produce sub-populations were chosen and %CV was calculated at the national and region levels. The four variables are:

- Percent for rent/sale
- Percent seasonal
- Number of buildings with 5-9 units
- Number of buildings with 10-19 units

Table 2 presents the results of the %CV calculations for various sample sizes. By design, %CV values decrease as sample size grows. At the current national sample size (~65,000), the %CV values for the four selected sub-populations are near 5 percent. Therefore, it can reasonably be concluded that the current national sample size meets the goal of the survey design.

It is worth noting, however, that the current AHS-N sample size does not produce Census Regional-level estimates that meet the %CV of five percent criteria. In fact, the AHS-N sample size would need to be increase three-fold to 200,000 before Census Region-level estimates approach %CV of five percent.

As noted in the section on predetermined updates to the 2015 AHS sample, there will no longer be a distinction between national and metropolitan area samples. Supplemental oversampling will occur in specific metropolitan areas as determined necessary by HUD. Although out-year budgets are difficult to predict, HUD assumes a total sample size of 185,000 – 200,000 for the 2015 AHS and beyond.

Table 2. Percent CV's for Selected AHS Sub-Populations

	For Rent/Sale				Seasonal			
	32,000	**65,000**	**100,000**	**200,000**	**32,000**	**65,000**	**100,000**	**200,000**
National	7.6%	5.3%	4.4%	3.1%	7.3%	5.2%	4.2%	3.0%
Census Regions								
Northeast	18.1%	12.8%	10.5%	7.4%	14.9%	10.5%	8.6%	6.1%
Midwest	15.1%	10.7%	8.7%	6.2%	16.6%	11.7%	9.6%	6.8%
South	11.5%	8.1%	6.7%	4.7%	11.7%	8.3%	6.8%	4.8%
West	18.0%	12.8%	10.4%	7.4%	16.0%	11.3%	9.2%	6.5%
	Buildings with 5 to 9 Units				**Buildings with 10 to 19 Units**			
	32,000	**65,000**	**100,000**	**200,000**	**32,000**	**65,000**	**100,000**	**200,000**
National	6.2%	4.4%	3.6%	2.5%	6.6%	4.6%	3.8%	2.7%
Census Regions								
Northeast	13.5%	9.6%	7.8%	5.5%	15.2%	10.8%	8.8%	6.2%
Midwest	14.2%	10.1%	8.2%	5.8%	14.8%	10.4%	8.5%	6.0%
South	10.2%	7.2%	5.9%	4.2%	10.2%	7.2%	5.9%	4.2%
West	12.3%	8.7%	7.1%	5.0%	13.8%	9.7%	8.0%	5.6%

Four Sample Design Options for the 2015 AHS

As documented in Dajani (2011), several criteria may be considered when evaluating which sample design best suits current and future stakeholders of the AHS. These include the ability to calculate cross-sectional and longitudinal estimates, the change from the current design, adaptability to changing geography and definitions, respondent fatigue, loss of housing units due to attrition of units and an uncertain budget environment.

In general, purely longitudinal sample designs are best at calculating long-term longitudinal estimates and involve the least amount of change from the AHS 1985 design. Purely cross-sectional (one response) sample designs are best at calculating cross-sectional estimates, are adaptable to changing definitions and conditions, do not suffer from attrition, and are robust to updating geography. Rotating panels are best at calculating short-term longitudinal estimates. They offer reasonable adaptability to changing definitions and conditions, suffer minimal attrition, and update geography reasonably well.

Because of the wide variety of users of AHS data and to consider challenges raised by the current survey design, four survey design options are presented. Reliability and precision can be assessed by referring to Table 2 for the predicted percent coefficients of variation for differing samples sizes.

Option #1 - Full Single Panel: Total sample size of 205,000, all units allocated in one single longitudinal panel for a panel lifespan of up to 30 years. The metropolitan oversampling design is included in Appendix 1.

Year	Panel
2015	00
2017	00
2019	00
…	00
2039	00
2041	00
2043	00

Option #2 – Full Dual Panel: Total Sample Size of 205,000 units divided equally into two samples of 102,500 units. The 2 samples would be interviewed on alternating schedule, with four years between each interview, i.e., Sample 1 interviewed in 2015, 2019, 2023, and so on and Sample 2 interviewed in 2017, 2021, 2025, and so on. The metropolitan oversampling design is included in Appendix 1.

Year	Panel	Panel
2015	01	
2017		02
2019	01	
…		02
2039	01	
2041		02
2043	01	

Option #3 – Hybrid Single and Rotating Panels: Total Sample Size of 205,000 units allocated with 100,000 units in the longitudinal panel and 105,000 units divided equally (21,000 units) among five rotating samples. The metropolitan oversampling design is included in Appendix 2.

Year				P	A	N	E	L	S			
2015	00	01	02	03	04	05						
2017	00		02	03	04	05	06					
2019	00			03	04	05	06	07				
2021	00				04	05	06	07	08			
2023	00					05	06	07	08	09		
2025	00						06	07	08	09	10	
2027	00							07	08	09	10	11

Option #4 – Rotating Panels: Total Sample Size of 205,000 units divided equally into five rotating panels of 41,000 units. The metropolitan oversampling design is included in Appendix 1.

Year			P	A	N	E	L	S			
2015	01	02	03	04	05						
2017		02	03	04	05	06					
2019			03	04	05	06	07				
2021				04	05	06	07	08			
2023					05	06	07	08	09		
2025						06	07	08	09	10	
2027							07	08	09	10	11

Sample Design Options Discussion: Attrition Rates and Sample Design Options

The evaluation of attrition rates in AHS showed that attrition rate increased considerably after 2005, or 20 years in sample. Depending on your view of what is an appropriate attrition metric and acceptable attrition rate, you may find the attrition rate unacceptable at a time prior to 20 years.

Sample design options 1, 2, and the 00 panel of option 3 assume that housing units remain in sample for up to 30 years. It may be reasonably expected that the trajectory of attrition rates in these two options will be similar to the trajectory of attrition rates found in the current AHS

sample, although the assumption cannot be confirmed until a future time. Option 2 may result in a smaller attrition rate because housing units are only sampled once every four years. It should be noted that options 1 and 2 could have a sample period of less than 30 years.

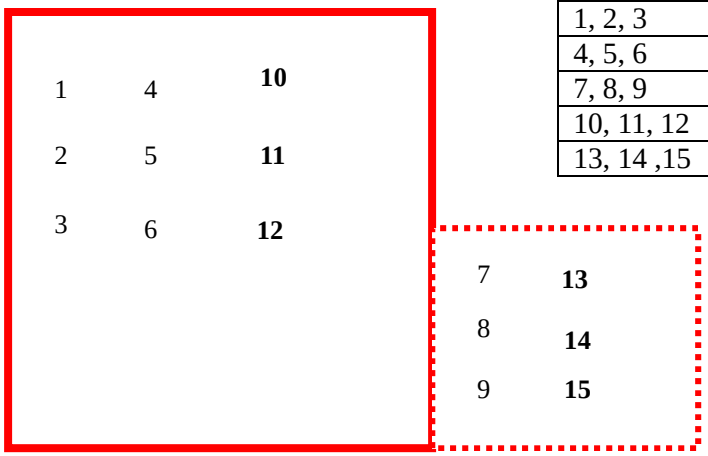
Sample design option 4 and the rotating panels in option 3 are in sample for only 10 years. The attrition rate analysis shows the attrition rates after 10 years to be 1 – 20 percent, depending on how attrition is measured. Attrition rates below 20 percent may be acceptable to most AHS users.

Sample Design Options Discussion: Geographic Identifiers

Unfortunately, each of the sample design options will suffer from disclosure issues caused by the inclusion of geographic identifiers. However, one strategy for dealing with disclosure issues is to use a technique known as "salting." Salting is the deliberate introduction or removal of sample units in order to introduce inaccurate results. In the case of the AHS, the salting technique could be used to remove observations that result in disclosure, as well as a random selection of observations that do not result in disclosure, and replacing them with an equivalent number of observations within and outside of the new geography.

Figure 2 shows how salting may be used to mitigate disclosure issues. In this example, observations 1 through 9 are part of the original sample, with 1 through 6 being inside a metropolitan area (denoted by the solid red box) and observations 7 through 9 being outside of the metropolitan area. Suppose the 2010 metropolitan area (denoted by solid red box) is expanded for 2020 by adding a new county (denoted by dashed red box). Observations 1 through 3 are kept in the sample. Observations 4 through 6 are randomly chosen to be deleted and replaced with observations 10 through 12. Observations 7 through 9 are deleted and replaced with observations 13 through 15. In this example it is impossible to determine which observations are in the new part of the metropolitan area.

Figure 2. Example of Salting Technique



Sample Point	Metro 2010	Metro 2020
1, 2, 3	Yes	Yes
4, 5, 6	Yes	Deleted
7, 8, 9	No	Deleted
10, 11, 12	------	Yes
13, 14 ,15	------	Yes

It is important to note that the salting technique may not work well when more than two geographic identifiers are present. In the example above, Metro 2010 and Metro 2020 are

identified. Suppose a third geographic identifier, County, was included. If the 2010 metropolitan area was composed of one county, and the county being added for the 2020 metropolitan area was part of another group of counties, then the mere inclusion of the county identifier on the PUF would produce a disclosure violation.

To understand this, consider Table 3, which is a modified version of the table from Figure 2, with an additional column for a county identifier. In this example, observations 13 through 15 are exactly identified as being in the new metro area and as part of county group 45. Anyone with a list of changes to the metropolitan areas between 2010 and 2020 could immediately determine that exact county where observations 13 through 15 reside because it has been disclosed that observations 13 through 15 belong to county group 45 and to a metropolitan area. This is a disclosure violation.

While salting can mitigate the disclosure problems of changing geography, it does have the side effect of reducing the longitudinal sample. In the above example, only observations 1 through 3 would be present in succeeding surveys. Observations 10 through 15 could not be used for longitudinal analysis of changes since the previous survey, although they could be used for such analysis in future surveys.

Table 3. Illustration of Disclosure with Three Geographic Identifiers

Sample Point	Metro 2010	Metro 2020	County
1, 2, 3	Yes	Yes	County 1
4, 5, 6	Yes	Deleted	County 1
7, 8, 9	No	Deleted	County Group 45
10, 11, 12	------	Yes	County 1
13, 14 ,15	------	Yes	County Group 45

Generally speaking, none of the sample designs will alleviate disclosure issues. Disclosure issues are only alleviated if geographies selected at the beginning of the sample remain fixed in vintage and no new geography types are added.

Sample Design Options Discussion: Maintaining a Representative Sample

Census has taken great care to ensure that the current AHS sample is representative of the current housing universe. Their strategy includes selecting new construction into the sample as well as a robust weighting scheme that benchmarks certain variables to other data sources. Furthermore, it can reasonably be concluded that analysis comparing the 2007 ACS to the 2007 AHS does not reveal bias caused by an out-of-date AHS sample.

The sampling strategies for the various design options are very similar, with the rotating panel design being slightly superior because sample is replaced at a greater rate and with benchmarks to current estimates. In addition, Census will continue to utilize a weighting strategy benchmarked to the most recent data sources available to ensure the sample continues to be representative of the housing universe.

Selecting New Sample for the Full Panel Designs (Options 1, 2, and the 00 Panel for Option 3)

The initial sample will be selected from the Master Address File at one time and new growth will be added.[9] The initial sample will be benchmarked to the 2010 Decennial Census and will be based on the following stratifications:

- Geography
- Tenure
- Type of structure
- Number of rooms[10]

New growth units added to the sample will be stratified only on geographic variables (county, tract, block, and zip code). It is expected than geographic stratification is sufficient to be representative of the types of housing units built since the last new growth sample was selected.

Selecting Sample in a Rotating Panel Design (Options 3 and 4)

The initial sample will be selected in the same manner as the initial sample in the single panel design (benchmarked to the 2010 decennial census). The differences in the sampling occur during subsequent sample years.

In subsequent years, the newly selected sample will be benchmarked to the 2010 census in terms of the type of housing unit (manufactured home, one unit structures, or two or more unit structures) and tenure (owned, rented, or vacant). These would then be benchmarked to the most current ACS data available in terms of home value and rent.

Sample Design Options Discussion: Achieving an Appropriate Sample Size in an Uncertain Budget Environment

A discussion of AHS sample design options must include how those options fare in response to an uncertain budget environment. In this context, the rotating panel design is superior to a single panel.

In the full panel designs (single or dual), the sample size is chosen up front. In the best case, the initial survey starts with a small sample size, then subsequently increases if budgets permit additional sample to be added to the current sample. Of course, if budgets were to decrease in subsequent years, the sample would have to be reduced. In fact, this has been the history of the current AHS-MS samples, where variations in budgets have resulted in both reductions and expansions to the number of AHS-MS conducted during a survey cycle. The net result has been an inconsistent pattern of AHS-MS surveys and limited longitudinality.

[9] Also, every two years, the subsidized housing sample will be refreshed to include the most current units.
[10] The MAF does not identify number of rooms. This measure will be based on a proxy measure using average home values and rents determined from the 5-year ACS data.

In the rotating panel designs, budget variations could be accommodated simply through reductions to the sample size for the upcoming panel. Consider Option 4 and that, for 2015, each panel costs $10 million dollars for 40,000 observations (for a total of 200,000 observations). If the 2015 budget permitted $50 million in the initial year, all five panels could be completed. Suppose that in the next survey cycle (2017), the budget was reduced by $5 million. As previously planned in the panel design, panels 2, 3, 4, and 5 could be completed for $40 million. Panel 6, the new panel, could be scaled back to include only 20,000 observations and $5 million, for a total survey cost of $45 million. If budgets were to return to 2015 levels for 2019, panel 7 could be increased to 60,000 observations, bringing the total observation back to 200,000. In all these cases, the longitudanality of the panels are maintained for ten years. Furthermore, the decrease in observations for 2017 is shared across the entire survey, as opposed to specific metro areas.

Conclusion

This white paper has presented four key issues concerning the current AHS sample design and presented four options for the 2015 AHS sample design. It is through discussion of these issues that AHS will continue to well serve the U.S. housing statistics community moving forward.

Acknowledgements

The authors thank Shawn Bucholtz, Dav Vandenbroucke, Carolyn Lynch, Wendy Chi, and Kurt Usowski for initiating the suggestions that motivated this paper. The authors also thank Joe Huesman of the Demographic Surveys Division and Tamara Cole of the Social, Economic, and Housing Statistics Division at the U.S. Census Bureau for their careful review and helpful comments on this paper.

References

Dajani, Aref N. (2011), “Redesigning the 2015 AHS Sample”, Handout and slides from presentation on April 6, 2011 to the National Association of Home Builders, U.S. Census Bureau.

Killion, Ruth Ann (2011), “Plan for the 2010 Sample Redesign of the AHSs”, Unpublished Memorandum to Distribution on February 2, 2011, U.S. Census Bureau.

Schwartz, Mary B. (2011), “2007 American Community Survey: A Comparison to Selected Housing and Financial Characteristics for the United State from the 2007 American Housing Survey”, U.S. Census Bureau.

Appendix 1. Metropolitan Oversampling Strategies for Proposed Full Panel Sample Designs

Starting in 2015, there will no longer be separate national (AHS-N) and metropolitan area (AHS-MS) surveys. A representative national sample will be surveyed during each survey. The national sample will be supplemented with additional "oversampling" cases in each of 60 metropolitan areas. Metropolitan oversample cases in 30 of the 60 metropolitan areas (Group A) will be surveyed in 2015 and in every other survey (2019, 2023, and so on). Metropolitan oversample cases in the other 30 of the 60 metropolitan areas (Group B) will be surveyed in 2017 and in every other year (2021, 2015, and so on).

The following tables show how the oversample will be incorporated into each potential sample design.

Table A1: Full Single Panel with Metropolitan Oversampling

Year	Panel
2015	National – 00 Metropolitan Oversample – Group A
2017	National – 00 Metropolitan Oversample – Group B
2019	National – 00 Metropolitan Oversample – Group A
…	
2039	National – 00 Metropolitan Oversample – Group A
2041	National – 00 Metropolitan Oversample – Group B
2043	National – 00 Metropolitan Oversample – Group A

Appendix 1. Metropolitan Oversampling Strategies for Each Sampling Design (continued)

Table A2: Full Dual Panel with Metropolitan Oversampling

Year	Panel	Panel
2015	National – 01 Metropolitan Oversample – Group 1	
2017		National – 02 Metropolitan Oversample – Group B
2019	National – 01 Metropolitan Oversample – Group A	
…		National – 02 Metropolitan Oversample – Group B
2039	National – 01 Metropolitan Oversample – Group A	
2041		National – 02 Metropolitan Oversample – Group B
2043	National – 01 Metropolitan Oversample – Group A	

Appendix 2. Metropolitan Oversampling Strategies for in a Rotating Panel Design

Year	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5	Panel 6	Panel 7	Panel 8	Panel 9
2015	National-1 MetOvr Grp A-1 ~~MetOvr Grp B-1~~	National-2 MetOvr Grp A-2 ~~MetOvr Grp B-2~~	National-3 MetOvr Grp A-3 ~~MetOvr Grp B-3~~	National-4 MetOvr Grp A-4 ~~MetOvr Grp B-4~~	National-5 MetOvr Grp A-5 ~~MetOvr Grp B-5~~				
2017		National-2 ~~MetOvr Grp A-2~~ MetOvr Grp B-2	National-3 ~~MetOvr Grp A-3~~ MetOvr Grp B-3	National-4 ~~MetOvr Grp A-4~~ MetOvr Grp B-4	National-5 ~~MetOvr Grp A-5~~ MetOvr Grp B-5	National-6 ~~MetOvr Grp A-6~~ MetOvr Grp B-1			
2019			National-3 MetOvr Grp A-3 ~~MetOvr Grp B-3~~	National-4 MetOvr Grp A-4 ~~MetOvr Grp B-4~~	National-5 MetOvr Grp A-5 ~~MetOvr Grp B-5~~	National-6 MetOvr Grp A-6 ~~MetOvr Grp B-1~~	National-7 MetOvr Grp A-2 ~~MetOvr Grp B-2~~		
2021				National-4 ~~MetOvr Grp A-4~~ MetOvr Grp B-4	National-5 ~~MetOvr Grp A-5~~ MetOvr Grp B-5	National-6 ~~MetOvr Grp A-6~~ MetOvr Grp B-1	National-7 ~~MetOvr Grp A-2~~ MetOvr Grp B-2	National-8 ~~MetOvr Grp A-3~~ MetOvr Grp B-3	
2023					National-5 MetOvr Grp A-5 ~~MetOvr Grp B-5~~	National-6 MetOvr Grp A-6 ~~MetOvr Grp B-1~~	National-7 MetOvr Grp A-2 ~~MetOvr Grp B-2~~	National-8 MetOvr Grp A-3 ~~MetOvr Grp B-3~~	National-9 MetOvr Grp A-4 ~~MetOvr Grp B-4~~
…….									

MetOvr Grp A = first 30 metros, MetOver Grp B = other 30 metros
~~MetOvr Grp A~~ denotes that the metro group within this panel is not surveyed in this year

Assumptions:

- Each metro oversample group would be interviewed 5 times over 20 years.

- A total sample size of 325,000 cases, comprised of:
  - A target national sample size of 85,000 cases, surveyed every year
  - A target metropolitan oversample sample size of 240,000 cases, with only half (120,000 cases) surveyed each survey year

  Therefore, only 205,000 of the 325,000 total sample cases are surveyed in each year.

- Each panel has a total sample size of 65,000 cases, comprised of:
  - National (17,000 cases)
  - metropolitan oversample for group A (24,000 cases)
  - metropolitan oversample for group B (24,000 cases)

  Therefore, only 41,000 of the 65,000 total panel sample cases are surveyed in each year.