Measuring Attrition in Long-Term Longitudinal Surveys

Aref N. Dajani[1], Shawn J. Bucholtz[2], and Toni M. Warner[3]

Overview

The American Housing Survey (AHS) is a biennial (every two years) 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, housing unit attrition, and oversampling of selected populations[4]. 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 housing units across time. The current sample is scheduled for final interviews in 2013. A new sample will be drawn for the 2015 AHS.

In addition to a national sample, the AHS maintains an oversample of targeted metropolitan areas that are interviewed less frequently. As the nature of this paper is to discuss AHS attrition from 1985 to the present, the scope of this paper will focus on the national sample.

Maintaining the AHS longitudinal sample for nearly thirty years has presented many challenges, including attrition of housing units, respondent burden, and changes in geography definitions requiring disclosure avoidance. 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.

An important performance measure for all surveys is the *response rate*. For the AHS, the response rate for a given survey year is the percent of eligible cases for which an interview was completed. The most frequent use of the AHS national sample is to produce cross-sectional estimates of variables of interest for a given year. A high non-response rate reduces the statistical precision of cross-sectional estimates, mostly because non-response reduces sample size and that increases variance.

Because the AHS national sample is longitudinal, the sample's *attrition rate* is also an important performance measure. For longitudinal surveys, including the AHS, attrition is defined as a reduction or decrease in size, number, or strength of the sample, whether permanently or intermittently. Attrition in a longitudinal survey reduces the statistical precision of *both* cross-sectional and longitudinal estimates and may result in sample bias.

Attrition can come in many forms. It is important to define attrition types for a survey before evaluating the performance of a survey. Generally, care should be taken to separate reasons for attrition into those that are not related to the survey design and those that are or may be related to the survey design.

---

[1] Longitudinal Surveys Branch, Demographic Statistical Methods Division, U.S. Census Bureau
[2] Office of Economic Affairs, Office of Policy Development and Research, U.S. Department of Housing and Urban Development
[3] Construction Programs Methodology Branch, Manufacturing and Construction Division, U.S. Census Bureau
[4] This will include subsidized housing for 2015 and may include oversampling certain metropolitan areas.

The American Housing Survey follows housing units over time by interviewing the households that occupy those housing units. Each year, a small percentage of the overall housing stock (and the AHS sample) is lost to demolition or disaster, so the AHS adds new construction to the sample for each survey. Historically, the AHS has been subject to budgetary fluctuations that reduce sample size in a particular year. Furthermore, changing priorities have resulted in sample additions for populations of special interest. The taxonomy includes four general types of attrition that will be defined later: permanent, administrative, construction-conversion-prohibited occupancy, and persistent non-response. They provide a useful framework for defining the taxonomy of attrition unique to the longitudinal AHS national sample.

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 nearly thirty-year panel, with final interviews scheduled for 2013. This sample has been supplemented by new housing 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 housing unit structure is of primary interest to the AHS. When a householder in an AHS housing unit moves, the structure remains in sample. The next AHS survey interviews the in-movers to the housing unit. Out-movers from the housing units are not interviewed.

The AHS longitudinal sample 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 to the end of the period of interest.

Measuring Non-response

There are multiple ways to measure non-response in a longitudinal survey, including an unweighted non-response rate or a weighted non-response rate. An unweighted non-response rate indicates the level of operational success that field representatives achieve in completing interviews. As unweighted responses usually cannot be projected to a larger population, it is preferable to refrain from conducting significance testing on statistics calculated from unweighted data. A weighted response rate gives an indication on how representative a sample remains with respect to the population to which the sample infers. The final weight typically incorporates response, so one uses the base weight, inversely correlated with the probability of selection. Statistical inference can be conducted on weighted non-response rates; hypotheses can be tested.

In addition, non-response can be calculated for those not interviewed or for those units where an interview is not attempted. For longitudinal surveys, non-response can be calculated cross-sectionally; i.e., for a specific time period, for consecutive time periods, for discrete pairs of time periods, or across all time periods where a unit is in sample.

Sample Frame, Sample Selection, and Interview Eligibility

It is necessary to understand how the AHS longitudinal sampling frame is structured. The *cumulative sample* includes all housing units (often referred to as cases) that have been visited since 1985.

As of the 2009 AHS, 102,594 cases are part of the cumulative sample; i.e., this is the total you would get if you combined each Public Use File from 1985 through 2009. These units were sampled at some point between 1985 and 2009. Of this total, 62,163 housing units were selected for sample in 2009. The 39% loss was due to prior demolished units, prior oversamples that are no longer of interest, and budgetary restrictions. Of those units, 59,555 housing units were selected for interview in 2009, with the resulting 3% loss due to newly demolished units or units under construction. Among those, there were 53,488 completed interviews. Therefore, the unweighted non-response rate is (59,555-53,488)/59,555 or 10.2%.

Permanently Ineligible Attrition (Prior or Current Determination of AHS Type C)

Housing units can be permanently removed from the housing stock for reasons that are unrelated to the survey, including economic reasons (demolition or relocation of housing unit) or disaster loss. These units are deemed permanently ineligible for the current AHS national sample. On average the AHS experiences permanent loss of housing units of **2.0 percent per survey year**. However, there is nothing inherent in the design of the AHS that causes this type of *permanent* attrition. The AHS addresses permanent attrition of housing units by adding new construction to the sample during each survey. This helps ensure that the sample remains representative of the housing stock. Also, it must be noted that this type of attrition is important to the AHS because it serves as a measure of housing stock loss.

Administratively Ineligible Attrition

An important characteristic of the AHS national sample is that once a housing unit is added to the sample frame, it is not deemed permanently ineligible for interview for any reason other than those discussed in the permanent attrition section. However, there are *administrative* reasons why housing units may be deemed ineligible for the sample for a given year, including budgetary reductions and special interest samples. Like many federal programs, the AHS has been subject to budgetary fluctuations throughout its history, and this has resulted in portions of the sample not being interviewed for a given year. For instance, eight percent of the 2005 sample was not interviewed due to budgetary restrictions.

Additional housing units had been added to the AHS national sample at various times for special purposes, including, metropolitan area oversampling, manufactured housing coverage improvement, and HUD-assisted renters oversampling. While these housing units have not been removed from the sample, many have been interviewed intermittently.

Administrative attrition is an unfortunate but common characteristic of many surveys. However, it is not related to survey design. That being said, HUD and the Census Bureau can do a better job of identifying specific reasons for administrative attrition in the PUF so analysts can better understand the historical nature of administrative attrition.

Construction\Conversion\Prohibited Occupancy Attrition (AHS Type B)

At the beginning of a survey cycle, new construction is added to the AHS sample. The Census Bureau selects new construction by determining where building permits have been issued. In some instances the unit is not habitable because construction has not started or is not complete at the time of survey. Even though these cases are eligible for the sample, they are deemed ineligible for an interview for the current year. However, they are eligible to be included in the sample in subsequent years.

In other instances, a housing unit is converted to an institutional unit or occupancy is either determined to be prohibited or not possible because the housing unit's interior is exposed to the elements. Similar to incomplete new construction, these housing units are deemed ineligible for an interview for the current survey but eligible for the sample in subsequent years.

AHS designates these housing units as "Type B" non-interviews. This type of attrition averages approximately **2.9 percent per survey year** and it is an important measure for understanding temporary changes to the housing stock, especially in light of housing market fluctuations. This type of attrition is unavoidable for housing unit surveys, but is not likely correlated with the survey design.

Further analysis is required to determine if the probability of a housing units being Type B for a given survey is significantly increased if it was previously determined to be a Type B. If this is the case, it may be beneficial for the AHS to deem these units permanently ineligible and replace them with additional sample.

Persistent Non-response Attrition (Type A non-interview)

The discussion up to this point has detailed three reasons why housing units are ineligible for the sample (permanent or administratively), or eligible for the sample, but ineligible for an interview (construction/conversion/prohibited occupancy). Although these types of attrition are unfortunate, there is little reason to believe that the design of the survey contributes to their existence or magnitude.

Attention is now focused on housing units where an interview was attempted but was not successful. The AHS designates these housing units as "Type A" non-interviews (referred to hereafter as non-response) for the survey year. However, they are eligible to be included in the sample in subsequent years. The reasons for Type A non-response are sub-classified in AHS and include:

- housing unit could not be located
- householder could not be located
- householder refused to participate
- language barriers prevented interview from taking place.

Virtually all surveys suffer from some level of non-response. Quite often, the most desirable type of non-response is non-response that is random across the population, meaning it is not correlated with survey questions or any particular sub-population. In cross-sectional surveys, random non-response means a smaller sample, which reduces the

statistical precision of the estimates. However, random non-response does not cause population estimates made from the sample to become biased.

It can reasonably be expected that the AHS will suffer from non-response during any particular survey. However, because the AHS is a longitudinal survey, non-response in one survey year could be related to non-response in a previous survey year; this is *persistent non-response attrition*. One reason may be that a householder refuses to participate in the current and future surveys because the previous survey was too long or the respondent feels the survey is administered too frequently (respondent fatigue). Another reason is that respondent may not be home during the times of the year for which the enumerator visits to administer the survey, and this happens each survey year. Still another reason is that the householder refuses to participate in any survey. Regardless of the reasons, if the non-respondent remains in the housing unit and the housing unit remains in the survey, the impact of persistent non-response attrition is a reduction in the longitudinal sample size beyond what might be expected due to random non-response, where current non-response is unrelated to prior survey nonresponse.

There are equally valid reasons why non-response rates for housing units that remain in the survey might be *lower* than what would be expected by random chance. For instance, if the survey is well designed and relatively easy for the respondent to complete, their prior experience may have been positive, so they may be likely to complete the survey in a subsequent year.

It is worth noting that households enter and exit housing units. As such, persistent non-response for a particular housing unit may end if a new household moves into the housing unit.

Although the reasons for negative and positive influences on non-response rates are important, what matters is the impact of non-response rates on the overall panel size. The following three subsections provide outcome measures for persistent non-response attrition.

Persistent non-response attrition: single year non-response rates

Most current users of the AHS create cross-sectional estimates for a single year and are mainly concerned about the single year non-response rate. Measuring the non-response rate for the entire sample and comparing it to the subset of cases that have been in the AHS national sample since 1985 can help determine if the length of time in the survey has an impact on non-response rates.

Table 1 presents the AHS Type A unweighted and weighted non-response rates for all interviewed housing units selected for interview for each year. Table 1 shows non-response rates were stable at 4 percent between 1985 and 1993, increased to 10% in 1997 and have remained stable since. This pattern is consistent with other surveys which experienced non-response rate increases in the late 1990's (Atrostic, et al., 2001). It is also important to note that the unweighted and weighted non-response rates, when different, are not vastly different, indicating that the patterns of non-response are not very different across strata.

Table 1. Single year non-response rates (standard error in parentheses) for all interviewed housing units.

Survey Year	Unweighted	Weighted (se)
1985	4.2%	3.9% (0.13%)
1987	3.2%	3.2% (0.13%)
1989	4.2%	4.1% (0.12%)
1991	4.4%	4.5% (0.11%)
1993	4.1%	4.3% (0.12%)
1995	7.5%	7.1% (0.13%)
1997	9.9%	9.8% (0.17%)
1999	9.9%	9.2% (0.15%)
2001	10.0%	9.9% (0.16%)
2003	9.2%	8.7% (0.14%)
2005	10.8%	10.9% (0.17%)
2007	12.4%	12.4% (0.17%)
2009	10.2%	10.5% (0.16%)

Table 2 presents the AHS Type A unweighted and weighted non-response rates for a subset of housing units: those selected for interview for each year that were part of the original sample selected in 1985. A comparison of Table 1 and Table 2 reveals a similar pattern of non-response rate increases. This similarity suggests that housing units in the sample since 1985 are no more likely to have higher non-response rates than housing units added to the sample in later years. Combining this finding with the known increase in non-response rates in other surveys, it is reasonable to conclude that response fatigue does not appear to be a factor in non-response rates.

Admittedly, this is an imperfect measure of non-response attrition due to response fatigue because it does not measure the non-response rate for *households* that have been in the survey for multiple years. While this type of analysis is certainly possible for the AHS, it is beyond the scope of this analysis. What matters, however, is the impact of persistent non-response attrition, and the magnitude does not appear very high.

Table 2. Single year non-response rates (standard error in parentheses) for those units sampled in 1985

Survey Year	Unweighted	Weighted (se)
1985	4.2%	3.9% (0.13%)
1987	3.1%	3.3% (0.13%)
1989	4.3%	4.1% (0.13%)
1991	4.1%	4.5% (0.11%)
1993	3.9%	4.2% (0.12%)
1995	7.5%	7.3% (0.15%)
1997	9.6%	9.7% (0.18%)
1999	10.2%	9.1% (0.16%)
2001	9.6%	10.0% (0.19%)
2003	9.2%	8.7% (0.16%)
2005	10.7%	10.8% (0.20%)
2007	12.1%	12.5% (0.21%)
2009	10.7%	10.6% (0.20%)

Persistent non-response attrition: consecutive years non-response rates

One of the popular uses of the longitudinal structure of the AHS is to measure near-term changes in housing units by using data from consecutive surveys. An example would include measuring the income difference between households who were interviewed in 2007 and 2009. Another example is the Components of Inventory Change (CINCH) analysis to determine housing stock changes (Eggers, 2010). For near-term changes it is important to have completed interviews for a pair of consecutive years, but it is not necessary to have completed interviews for all years.

Table 3 presents the non-response rates for pairs of consecutive years. Table 3 shows that non-response rates for pairs of consecutive years have been between 15 and 18 percent since 1997. Most of the joint non-response rates can be attributed to non-response in one of the survey years. Also, the pattern in Table 3 mimics the pattern in Table 2 where non-response rates began to increase in 1995, but have been steady since then, with a minor uptick in 2007.

Table 3. Unweighted and weighted non-response rates for housing units selected for interview in consecutive pairs of years (standard error in parentheses)

Survey Years	Unweighted	Weighted (se)
1985 & 1987	5.9%	5.9% (0.20%)
1987 & 1989	6.0%	6.1% (0.20%)
1989 & 1991	7.0%	7.0% (0.20%)
1991 & 1993	7.1%	7.1% (0.13%)
1993 & 1995	9.5%	9.4% (0.17%)
1995 & 1997	13.7%	13.2% (0.18%)
1997 & 1999	15.4%	15.2% (0.20%)
1999 & 2001	15.6%	15.0% (0.18%)
2001 & 2003	15.2%	14.8% (0.20%)
2003 & 2005	15.7%	15.2% (0.18%)
2005 & 2007	17.9%	17.9% (0.21%)
2007 & 2009	17.8%	17.9% (0.20%)

Persistent non-response attrition: intermittent attrition for the panel

A small number of AHS users require the longitudinal panel to have a completed interview for most or all years for which the housing unit has been in the sample. These AHS users are typically measuring spells of activity or event-history analysis, such as "underwater" mortgage status over time, where the loan due on a house is greater than its market value. AHS user's ability to conduct longitudinal analysis with the full panel may be impacted by missing data in one or more years. Non-response (missing interviews) for one or more than one year is known as intermittent attrition.

Table 4 present cumulative and average intermittent attrition rates for the 60,932 housing units that entered the sample in 1985 or later and remained in the sample through 2009, including 35,053 housing units that entered the sample in 1985.

Table 4 reveals several important intermittent attrition rate outcomes for the AHS:

- Forty-seven percent of the housing units that have been in the survey since 1985 have non-responses. Therefore, 53 percent have no non-responses whatsoever.

- Of the housing units that have been in the sample for 13 surveys (since 1985), only 16 percent have 3 or more Type A non-responses. Therefore, 84 percent of housing units in the sample since 1985 have at least eleven completed interviews.

- The average number of Type A non-responses, which could be expected to rise as the longitudinal sample grows longer, appears to level off at 1.2 – 1.3 missing interviews, after being in the sample for at least eight interviews.

Table 4. Intermittent non-response attrition rates for selected survey lengths

Number of survey years in sample	Percent with 1 or more Type A non-responses	Percent with 2 or more Type A non-responses	Percent with 3 or more Type A non-responses	Average number of Type A non-responses
13 (since 1985)	47%	26%	16%	1.2
12 (since 1987)	52%	29%	17%	1.3
11 (since 1991)	50%	28%	16%	1.2
10 (since 1993)	48%	27%	15%	1.2
9 (since 1995)	44%	25%	13%	1.0
8 (since 1997)	44%	24%	15%	1.1
7 (since 1999)	42%	20%	9%	0.8

Discussion and Future Directions

Understanding attrition is necessary for judging the performance of a longitudinal survey. Perhaps most importantly, it is critical to determine why a particular type of attrition might occur and, if necessary, focus on the types of attrition that may be related to the design of the survey.

The AHS experiences several types of attrition. However, some of them are not related to the design of the survey. Rather, they can be said to "come with the territory" of administering a survey that evaluates a changing housing stock, in an environment that is subject to budget and priority fluctuations.

This attrition rate analysis evaluated the types of attrition unique to the AHS and focused on the types of attrition that may be, or likely are, related to survey design. The findings of this analysis suggest that attrition is not a serious problem for the AHS. Single year Type A non-response rates have stabilized at around ten percent.

In future research, we will determine the chance that a household remains a Type B household given they were deemed Type B in the previous interview. We will also look at attrition metrics as they may indicate data quality, as survey technology continues to evolve during the lifecycle of a long-term longitudinal survey.

Acknowledgements and Disclaimer

References

Atrostic, Bates, Burt and Silberstein, (2001). "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." Journal of

Official Statistics, Vol. 17, No. 2, pp. 209-226.

Eggers, Frederick J. (2010). "Components of Inventory Change." A report prepared for the Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC.

Johnson, Timothy. P. and Linda Owens (2003). "Survey Response Rate Reporting in the Professional Literature." Proceedings of the 2003 Conference of the American Association for Public Opinion Research, Nashville, TN.