

Impact of Dependent Interviewing on Consistency of Answers in the American Housing Survey

Katie Gustafson and Evan Brassell

American Housing Survey Branch, Social, Economic & Housing Statistics Division, U.S. Census Bureau. 4600 Silver Hill Road, Suitland, MD 20746.

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research and Policy Conference

This report is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Abstract

Dependent interviewing is the use of a respondent's answers from a previous wave in a longitudinal survey to modify question phrasing or skip items, with the goal of reducing respondent burden and improving data quality. The American Housing Survey is a longitudinal survey conducted every two years, in which dependent interviewing has been used since 1974, with the exception of waves for which a new sample was drawn. In 2015 a new sample was drawn, and for some questions that had been dependent interviewing questions in 2013 and earlier, notable differences were found between 2015 estimates and estimates from previous waves. Examination of multiple waves of responses indicated that dependent interviewing may exert an influence by reducing changes in response and reducing uninformative responses, effects which have also been observed in other surveys, and which often motivate use of dependent interviewing.

Introduction

The American Housing Survey (AHS), a longitudinal survey conducted by the U.S. Census Bureau and sponsored by the U.S. Department of Housing and Urban Development, is the most comprehensive housing survey in the U.S. and a prime data source for analyses of the nation's housing needs that inform federal housing policy. Rather than households, which are composed of people occupying a home, the sample is composed of the homes themselves, referred to here as housing units, whose occupants are visited or called every two years by trained interviewers. The sample that was interviewed beginning in 1985 was retired after the 2013 survey, and a new sample was drawn for 2015.

For the 1985-2013 sample, the survey had been simplified in various ways for returning respondents, using answers provided by them or by a previous occupant in an earlier wave of the survey. A number of questions were rephrased to include reminders of the respondent's previous wave answer, and some were skipped entirely. For example, if a housing unit was reported to be part of a condominium in the 2011 survey response, then in 2013 the respondent was asked, "Last time we recorded your home was part of a condominium. Is this information still correct?" Otherwise, the question was phrased: "Is this home part of a condominium?" If a homeowner provided the source of their down payment in 2011, then the down payment source was not asked of that household again in 2013.

This rephrasing or skipping of questions based on previous wave responses is called dependent interviewing. In the AHS, it is used in questions where the response is generally expected not to change (eg, number of stories in an apartment building), to change predictably (eg, respondent age), or to change infrequently (eg, whether the building containing the unit is part of an apartment complex). Dependent interviewing has been used in the American Housing Survey in every survey having a returning sample, with its use increasing considerably over the past two decades - from fewer than ten dependent interviewing questions in surveys through 1995, to 164 in 2013, with this increase driven by the switch from a paper instrument to a computerized instrument in 1997 (U.S. Census Bureau 2016a).

The new sample in 2015 motivated interest in effects dependent interviewing might have on consistency of respondents' answers, which could in turn influence item response distributions. For some items, notable differences were found between the 2013 and 2015 estimates even when accounting for changes in sample design, and dependent interviewing was posited as one possible contributor. This paper presents an overview of dependent interviewing, followed by a collection of evidence of two effects of dependent interviewing in the AHS: reduction in item uninformative response percentages and reduction in variability of a respondent's answers to a question across waves.

Overview of Dependent Interviewing

Principal aims of dependent interviewing (DI) are to reduce the burden on respondents of completing the survey and to improve data quality (Jäckle 2005). Research over the past 25 years points to three effects of dependent

interviewing on data quality: it reduces item nonresponse, reduces spurious change, and reduces spurious stability (Moore 2006, Jäckle 2009, Lynn, Jäckle, Jenkins and Sala 2005). For example, it has been shown to reduce underreporting of income sources and to decrease spurious change in occupation and industry of employment (Lynn, Jäckle, Jenkins and Sala 2005; Hill 1994). Its effects on data quality may be due to its streamlining of the interview process: It simplifies the response task and can decrease interview length, reducing both respondent and interviewer burden (U.S. Census Bureau 2016b; Weinberg 2002; Jäckle 2008; Sala, Uhrig and Lynn 2011).

Dependent interviewing could conceivably allow respondents to use “least effort strategies” such as supplying the same answer at each wave regardless of whether change has occurred (Jäckle 2009). By simplifying the response task, DI is meant to reduce such “satisficing” – providing minimally acceptable answers rather than optimal ones (Krosnick 2000). Rather than satisficing, the complexity of the response task may motivate erroneous confirmation of previous responses; for example, recipients of welfare benefits with a more intricate history of receipt were found to be more likely to agree with false previous information about their history (Eggs and Jäckle, 2015). In general, dependent interviewing questions that ask for a complex response, such as those on income sources, may be more prone to measurement error arising from underreporting of change.

One suggested way to reduce underreporting is to ask the DI question of a respondent who reported the characteristic of interest during any of the last k waves of the survey, not just the immediately previous one¹ (Lynn, Jäckle, Jenkins and Sala 2012). For example, in a $k=3$ design, if a housing unit was reported to have central air conditioning in wave 1, then in waves 2, 3 and 4 the respondent at the unit would be asked the DI question (“Previously we recorded that your home has central air conditioning. Is this information still correct?”; Figure 1). If this respondent answered “No” in waves 2, 3 and 4, then in wave 5 the independent (INDI) version (“Does your home have central air conditioning?”) would be asked.

Types of Dependent Interviewing

Dependent interviewing questions may be of two types: proactive or reactive (Lynn, Jäckle, Jenkins and Sala 2005). Proactive dependent interviewing uses responses from a previous wave of the survey or administrative data to change the routing or phrasing of questions for a respondent (Jäckle 2009). Two types of proactive DI are used in the AHS (U.S. Census Bureau 2016a):

- Do Not Re-Ask. Once a valid answer is obtained, the question is not asked in any future wave interview at that housing unit. The question on square footage of the housing unit in the 2017 survey is one example. Some Do Not Re-Ask questions are household-specific and so are re-asked of new occupants, such as questions on mortgage details. Do Not Re-Ask questions are asked at a housing unit if no interview was obtained at the unit in the previous wave.
- Verify. The respondent is asked to confirm no change since the previous survey. In 2013, for housing units inside a multiunit building, the number of stories in the building was asked as a Verify question: “Last time we recorded 3 stories in this building. Is this information still correct?”

Reactive dependent interviewing is a sort of edit check that occurs during the interview (Jäckle 2009). If the respondent’s answer significantly deviates from or contradicts a response to the same question in the previous wave, then a follow-up question asks for confirmation or clarification regarding that difference. Reactive DI can be used to decrease item nonresponse, where the respondent who answers “I don’t know” is reminded of their answer from the previous survey, and asked if there is any change.

¹ Lynn, Jäckle, Jenkins and Sala (2012) refer to this as an $n=i$ design. Here we use k rather than n .

Historically, reactive DI has been used for few questions in the AHS because of its potential to increase survey length and irritate respondents by questioning the correctness of information they have just provided. Two types of reactive DI questions have been used in the AHS (U.S. Census Bureau 2016a):

- **Probe.** The interviewer asks a follow-up question to confirm a reported change from the previous wave answer that is unexpected or large. In the 2011 survey, the question on the owner's estimate of the value of the home was a Probe question. If the reported home value was at least 35 percent higher or 35 percent lower than the value reported in 2009, then the respondent was asked: "Last time we interviewed your household, the value of the property at which you live was reported to be \$X. This time you reported the value to be \$Y. This represents a substantial change. What is the reason for the large difference?"
- **Bound.** A respondent is asked whether two infrequent events of the same type that were reported in the previous and current wave are the same event. In the Home Improvement section of the AHS, when a respondent reports a home repair project of the same type as a project reported by the household in the previous wave (eg, replacement of a roof in both 2015 and 2017), the respondent is asked if these are the same project. If they are, the interviewer does not ask for any further information about the project, and any record of that project is deleted from the current wave dataset during data processing.

The 2017 AHS contained two reactive DI questions: a Probe question on the type of structure containing the unit (single unit attached, single unit detached, mobile/manufactured home, multiunit building), which the interviewer often answered, and the Bound question referred to above on home improvement projects reported twice in consecutive waves.

When to Use Proactive Dependent Interviewing

Jäckle (2008) listed six situations in which proactive DI offers the most potential to increase efficiency and reduce burden (AHS examples are given for situations 1 through 4; situations 5 and 6 do not apply to the AHS):

1. **Skipping multiple follow-up questions.**
In 2013, the bulk of the mortgage section was skipped for homeowners who reported having only one mortgage in 2011, and having the same single mortgage in 2013.
2. **Accelerating respondent recall for difficult questions.**
In 2013, respondents who had reported their home water source and the number of units served by the water source in a previous wave were asked to verify that the information was still correct.
3. **Simplifying the set of answers the respondent selects from, eg, by asking a yes/no question rather than an open-ended question.**
The question on primary heating fuel asks the respondent to choose from a list of nine fuel types. The DI version of this question provides the previously reported fuel type and asks whether it is the same.
4. **For questions about relatively stable characteristics.**
For a housing unit previously reported to have a porch, the respondent is asked: "Last time we recorded that the house had a porch, deck, balcony, or patio. Is this information still correct?"
5. **For surveys with short intervals between interviews.**
6. **Skipping questions that are coded from free text responses post-fieldwork, reducing interview time and coding time.**

Effects of Dependent Interviewing in the AHS

Data from waves 2007-2015 was examined for evidence of whether two effects of dependent interviewing observed in other surveys – reducing item uninformative responses ("I don't know" or refusal to answer) and reducing change in answers across waves – occur in AHS data; a tour through the results of that examination follows.

Change in Estimates Obtained From Yes/No Questions

AHS estimates of the nationwide prevalence of a number of housing characteristics, eg, how many homes in the U.S. have a garage, come from responses to yes/no questions. For each national estimate derived from a yes/no question between 2007 and 2015, the percentage change in the estimate from the previous wave was obtained (wave-to-wave change in the estimated total number of housing units in the U.S. was not part of this calculation). Of the four waves examined, the most extreme overall drop in estimated prevalence was observed in the 2015 wave (Figure 2).

When the magnitude of the change was examined (Figure 3a), the absolute change in estimated prevalence for 2013-2015 was higher on average than that of 2007-2009, 2009-2011, and 2011-2013². This was compelling, but did not necessarily point to dependent interviewing being a factor.

Given that some DI questions are not re-asked in a later wave once an informative answer is obtained, and that questions selected to be DI questions are often about features of the home that change infrequently, it might be expected that dependent interviewing questions would show a lower absolute percentage change than INDI questions. Each set of yes/no questions occurring in waves 2007-2009, 2009-2011, and 2011-2013 was partitioned into DI and INDI questions, and their distributions of absolute percentage change values were compared within each wave (Figure 3b; in each wave, fewer than one third were DI questions). Any difference in medians between DI and INDI questions was not of practical significance (Table 1). The INDI questions appeared to have more extreme absolute percentage change values, evidenced by their wider spread (see Table 1 for dispersion estimates), which would be expected under the assumption that DI holds down variability in a housing unit's responses.

Result of Removing Dependent Interviewing from a Question

Some questions in the AHS are expected to have highly stable answers from wave to wave, because they address characteristics of the unit that change infrequently. These are asked as dependent interviewing questions. One such question is "What year was your home built?" which is asked in any of the following instances:

- When a housing unit is new to the sample
- If no year built value is recorded for a housing unit that has been surveyed previously, for example if the respondent has answered "I don't know" in every prior survey
- When the tenure of a housing unit changes from renter to owner, with the assumption that an owner would be more likely than a renter to know the year the structure was built
- If an interview was not completed for a housing unit in the previous wave, in which case no dependent interviewing questions are asked in the subsequent wave.

Once a year value is recorded for the housing unit, the question is not asked again.³

In 2009, this question was asked of about 30,000 housing units that already had a year built value stored for 2007, because the year built values from the 2007 survey had not been loaded into the instrument before the 2009 interviews were conducted. The 2007 values were compared to the values provided by respondents in 2009, to develop a picture of the stability of responses in the absence of dependent interviewing. The observations used in this comparison were housing units with a completed interview in 2009 having a non-imputed, unallocated and non-missing year built value in the 2007 response, and which did not have a change in tenure from renter to owner in 2009. Year built was a DI question starting in 1984, so an undetermined proportion of year built values from the

² Median for 2007-2009: 3.11%; 2009-2011: 5.40%; 2011-2013: 5.50%; 2013-2015: 22.69%.

³ Respondents who provided an answer in 2015 were asked this question in 2017. Dependency was removed from this question and from a small set of other questions in the 2017 survey in order to evaluate the impact of dependent interviewing, and because of constraints on time to test the instrument software.

2007 survey were provided before 2007. Since sampling is done on housing units rather than households or individuals, a previous household or a different individual may have provided the year built value stored for 2007.

Year built is a categorical question with response values spanning 1, 5 or 10 years (see Table 2 for a list of the year built categories in the 2007 survey). In 2009, 55 percent of respondents reported a value that did not match the value from 2007 (Table 3). If these categories are collapsed into decades, then this percentage is 45 percent. Over 10 percent of housing units had a missing or “Don’t know” value in their 2009 response. This percentage varies with the decade values given in 2007. Figure 4 shows the percentage of discrepant responses and the percentage of 2009 “Don’t know” responses for each decade built answer from 2007, where a discrepant response is one that in 2009 was not the same decade as the response provided in or before 2007.

Occupants of newer units, as defined by the year built value from 2007, appeared to be more likely to report a consistent year built value than those of older units⁴, and less likely to respond with “I don’t know”⁵. The housing units reported in 2007 to have been built in 1930-39 had the highest percentage discrepant (79%). For that group, the discrepant 2009 responses clustered around 1930-39, with the preceding two decades and the following two decades containing over 80% of the discrepant responses.

A similar pattern was observed with the other decade categories (Figure 5): the decade built value from 2007 was the most commonly reported value in 2009, and 2009 values clustered around the 2007 value, with 49.3 percent to 88.4 percent of 2009 values falling within one decade of the 2007 value (Table 4). On average, the 2009 decade built values matched the corresponding 2007 decade built values, but individual responses were quite variable. It may be that this question is difficult for respondents to answer consistently from wave to wave, simply because they may not actually know the answer.

While year built values can in some cases be obtained from administrative records, which would reduce within-respondent variability across waves, the high percentage of discrepant responses is worth noting generally when considering whether to skip a question in subsequent years once a response is obtained, particularly for quantities for which many respondents might only be able to provide an informed guess.

Increase in Percentage Uninformative More Likely in Absence of Dependent Interviewing

For any question, a respondent can refuse to answer, or can answer “I don’t know,” both of which are recorded and are considered uninformative responses. In 2015 the percentage of responses that were uninformative increased for the question that asks for the size of the unit in square feet, which is skipped for a household in future waves once an informative answer is obtained, and is re-asked if there is a change of households or if the previous wave interview was not completed (U.S. Census Bureau, 2016a). In each wave between 2007 and 2013 when unit square footage was a proactive DI Do Not Re-Ask question, not more than 13.0 percent of units had an uninformative response recorded⁶. In 2015 when unit square footage was not a dependent interviewing question because a new sample was surveyed, 22.4 percent of responses were uninformative⁷.

The bridge sample is a subset (n=6,000) from the 1985-2013 sample that was surveyed again in 2015, and was not included in 2015 estimates. Since there were only INDI questions in the 2015 survey, the 2015 responses from this group could be compared to their responses to the 2013 survey to see whether responses tended to change in the

⁴ Cochran-Armitage test of linear trend in percentage answers consistent between 2007 and 2009, and decade unit was built reported in 2007, p-value<0.0001.

⁵ Cochran-Armitage test of linear trend in percentage informative answers in 2009 and decade unit built reported in 2007, p-value<0.0001.

⁶ 2007: 9.9%; 2009: 10.1%; 2011: 12.9%; 2013: 11.9%. These are unweighted percentages.

⁷ For each test of no difference between proportion uninformative in 2015 and proportion uninformative in a previous wave (2007, 2009, 2011, 2013), p-value<0.0001.

absence of DI. Of the 6,000 housing units in the bridge sample, 3,908 had completed interviews in both 2013 and 2015 and were used for analysis. For that group, the percentage uninformative responses to the unit square footage question was 9.0 percent in 2013, and 25.3 percent in 2015⁸. This jump in uninformative response percentages for the same group over two waves motivated a wider examination using the bridge sample.

The bridge sample's uninformative response percentages were calculated for each housing unit-level question asked in both 2013 and 2015 having at least 25 responses in each wave ($n_{\text{questions}}=144$). Confidence intervals of the difference in matched pairs proportions were constructed for each question (95% Adjusted Wald CI (Agresti and Min 2005); Figure 6). Questions were put into one of three categories based on the bounds of their confidence interval: percentage uninformative significantly decreased, did not change significantly, or significantly increased, from 2013 to 2015 (Table 5). Evidence suggested an increased percentage uninformative in 2015 was more likely for questions that were dependent interviewing questions in 2013, compared to those that were not⁹. For questions that had been selected to be dependent interviewing questions (eg, those with answers expected to change infrequently), respondents seemed more likely to provide an uninformative answer when that question was not asked as a dependent interviewing question.

Stability of Responses Across Waves

Room Counts

AHS respondents are asked for the count of each of several types of rooms in the unit, eg, the number of bedrooms and number of bathrooms. In the 2009 and 2011 waves, these were INDI questions. In 2013, they were DI questions. The room counts from the national sample housing units having a completed interview in all three waves ($n=28,256$) were examined for evidence of whether a respondent's answers tended to change less frequently from wave to wave with DI questions than with INDI questions. Changes in responses to bedroom counts in the 2009-2011 waves of the AHS have been shown to be very rarely due to structural changes to the home, so responses to questions about home improvement projects were not examined to determine if room additions or removals had been reported (Eggers and Moumen 2013). Table 6 shows the percentage of responses with a change in room count between 2009 and 2011, and between 2011 and 2013, for each of the room types. Across room types, the percentage in 2013 was on average significantly smaller than in 2011¹⁰ (Figure 7), and within each room type, the difference was significant¹¹.

A relatively high percentage of answers were discrepant for number of dining rooms in the unit: 25.4 percent between 2009 and 2011, and 15.6 percent between 2011 and 2013. Over 90 percent of these discrepant answers were zero in one wave, and one in the next, or vice versa. The percentage of respondents switching from reporting zero dining rooms to reporting one or more appeared to be similar to the percentage switching from one or more to zero across adjacent waves in 2007-2011 (Figure 8), but the percentages were statistically significantly different¹².

The percentage switching from reporting one dining room or more in 2011 to zero in 2013 seemed particularly small at 3.7 percent (Table 7) compared to the 10.9 percent switching from zero in 2011 to one or more in 2013¹³. In 2013,

⁸ Test of no difference in proportion uninformative between 2013 and 2015, $p\text{-value}<0.0001$.

⁹ Cochran-Armitage test, $p\text{-value}<0.0001$ for hypothesis of independence versus alternative of linear association between a question being DI in 2013 and a significantly increased percentage uninformative in 2015, where a significant difference was one in which the 95% Adjusted Wald CI of the difference in percentage uninformative between 2013 and 2015 did not contain zero, for the set of respondents who answered the question in both waves.

¹⁰ One-sided Wilcoxon signed rank test of zero median difference in 2011 and 2013 percentage responses discrepant with previous wave, $p\text{-value}=0.0002$.

¹¹ McNemar's test of marginal homogeneity on a contingency table of 2011 answer discrepant with 2009 by 2013 answer discrepant with 2011, $p<0.0001$ for each room type.

¹² 2007-2009 $p\text{-value}<0.0001$, 2009-2011 $p\text{-value}=0.035$, McNemar's test of marginal homogeneity.

¹³ For test of no difference between 3.7 and 10.9 percent, $p\text{-value}<0.0001$.

the number of dining rooms was a DI question, but was only asked as a DI question of respondents who had reported at least one dining room in 2011. Those who had reported zero dining rooms in 2011 were asked the INDI version of the dining rooms question. This might explain why the percentage of respondents switching from one or more in 2011 to zero in 2013 was markedly smaller than in previous waves, and is supportive of the assertion that dependent interviewing may reduce variability in a unit's responses from wave to wave in the AHS. This also suggests that asking the DI version of a question only when the respondent's previous wave answer was greater than zero could result in upward bias of estimates, and merits further investigation.

Estimated Effect of Dependent Interviewing on Probability of Discrepant Answer to Home Fuel Questions

In order to quantify the influence of dependent interviewing on whether an answer would be discrepant with the previous wave answer while taking into account other factors, a longitudinal model (generalized linear mixed model, binary response, covariance structure of compound symmetry) was constructed of the probability that a respondent's answer to a question on home fuel type would change from the previous wave, based on whether the respondent was asked the DI form of the question, and using the responses from waves 2011, 2013 and 2015. The model controlled for whether the unit was owned or rented by its occupants, structure type, and the fuel question asked (fuel used for cooking, drying clothes, heating the home, or heating water). The model did not control for whether different households completed the survey for the housing unit in the two waves being compared, nor for whether a different respondent within the same household completed the survey. The phrasing of the questions did not change across the waves examined, but the answer categories for water heater fuel type and cooking fuel type changed in 2015, with the categories for kerosene, wood and coal being collapsed into the "Other" category. In the analysis, the three categories were retained in the 2011 and 2013 comparison, but were collapsed to "Other" for the 2013 to 2015 comparison. Because prediction was not so much of interest as quantifying a possible influence of DI while controlling for other factors, no model selection procedure was undertaken.

The data to train the model was a simple random sample ($n_{\text{train}}=1,890$) of the bridge sample respondents having answers for at least one fuel type question in either 2011 and 2013 or 2013 and 2015 ($n=3,151$). The data to test the model were the respondents in the bridge sample that had not been selected for the training set ($n_{\text{test}}=1,261$).

The model fit was considered adequate in light of its concordance index of 0.751 (95% CI: 0.733, 0.769; 0.5 would indicate results no better than a random guess). Setting the cutoff for predicted probability of discrepant answer to 0.08 (the approximate proportion of observations that were discrepant in the training dataset) yielded sensitivity of 0.701 and a correct classification rate of 0.685, for the test data.

The model estimated that the odds of a respondent supplying a discrepant answer was 41 times higher when a question was asked without dependent interviewing than with (95% CI of odds ratio: 25.5, 66.8) when controlling for the other factors, all of which were significant at the $\alpha=0.05$ level (Table 8). Rather than being an indication of an extreme influence of dependent interviewing, the large odds ratio estimate and its wide confidence interval signal sparse data for some combinations of outcome and explanatory variables (Greenland, Mansournia and Altman 2016). Treating wave of survey as a factor, over 60 percent of these combinations had a cell count less than 5. Since the responses from 2013 included in the analysis were only from those who had completed the 2011 survey, most responses were to DI versions of the questions in 2013 (98.7% in the training set). All responses in 2015 were to the INDI versions of the fuel questions, since the 2015 instrument had no DI questions. While the dataset allowed for examination of a subject's answers to both the DI and INDI versions of a question, the asymmetry in distribution within each wave of questions asked DI or INDI resulted in a dataset that precluded precise parameter estimates.

Conclusion

Evidence suggests an influence of dependent interviewing on AHS responses. It appears to dampen a housing unit's variability in answers to a question across waves, and reduce the proportion of responses to a question that are uninformative within a wave. Such effects of dependent interviewing have been observed in other surveys.

Future studies of effects of dependent interviewing in the AHS might include comparing responses against administrative records to look for evidence of bias in response distributions attributable to dependent interviewing. There is interest in determining whether some forms of dependent interviewing, such as Do Not Re-Ask, might introduce spurious stability across waves.

Analyses that include the 2017 data, which was not available at the time of writing, might provide a fuller picture of any impact of the 2015 redesign. A revisit of the model of probability of a discrepant answer could include training the model with a dataset with fewer sparse cell counts, in order to obtain a more precise estimate of the influence of dependent interviewing. Possible modulation of effects of dependent interviewing in the AHS by whether the response comes from a new household, or a different respondent in the same household, remains a compelling open question.

Acknowledgements

Significant and valuable contributions to discussions that were foundational to this paper were made by Sam Garrow and Annie Leung. Sam's analysis of room count variability was the basis for the analysis in the *Room Counts* section.

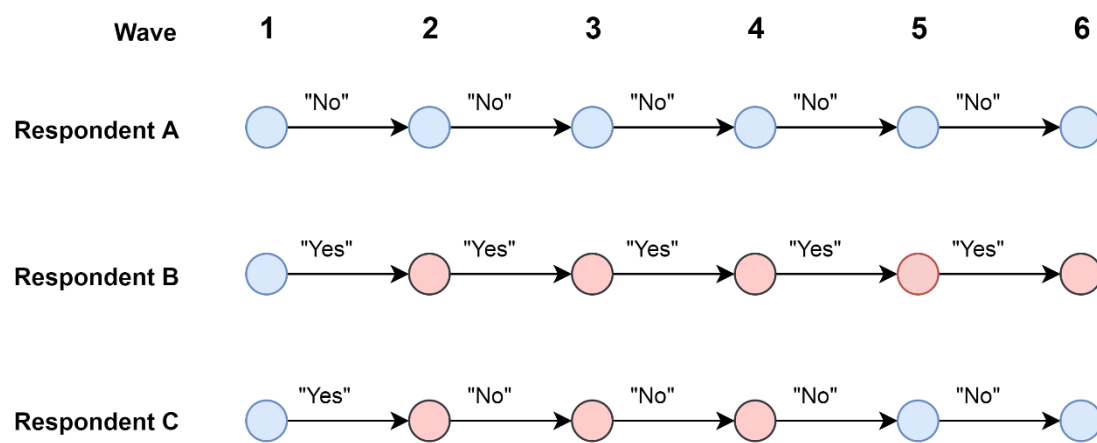
Sue Lord informed the authors of the existence of the year built re-ask that occurred in 2009. Lisa Aronhalt Ferrer directed the authors to the correct datasets for the year built analysis, and laid out the universe of responses for the comparison.

References

- Agresti, Alan and Yongyi Min. 2005. "Simple Improved Intervals for Comparing Matched Proportions," *Statistics in Medicine* 24:729-740.
- Eggers, Frederick and Fouad Moumen. 2013. "Between-Survey Changes in the Number of Bedrooms in a Unit: How Often? Why? Effect on Measures of Rental Affordability?" Paper prepared for the Office of Policy Development and Research, U.S. Department of Housing and Urban Development. Available at https://www.huduser.gov/portal//publications/pdf/AHS_NumberOfBedrooms.pdf.
- Eggs, Johannes and Annette Jäckle. 2015. "Dependent Interviewing and Sub-Optimal Responding," *Survey Research Methods* 9(1):15-29.
- Greenland, Sander, Mohammad Ali Mansournia, Douglas G. Altman. 2016. "Sparse Data Bias: A Problem Hiding in Plain Sight," *British Medical Journal* 2016;352:i1981.
- Hill, Daniel. 1994. "The Relative Empirical Validity of Dependent and Independent Data Collection in a Panel Survey," *Journal of Official Statistics* 10(4):359-380.
- Jäckle, Annette. 2005. "Does Dependent Interviewing Really Increase Efficiency and Reduce Respondent Burden?" *Institute for Social & Economic Research Working Papers* 2005-11.
- Jäckle, Annette. 2008. "Dependent Interviewing: Effects on Respondent Burden and Efficiency of Data Collection," *Journal of Official Statistics* 24(3):411-430.
- Jäckle, Annette. 2009. "Dependent Interviewing: A Framework and Application to Current Research," Pp. 93-111 in *Methodology of Longitudinal Surveys*, edited by Peter Lynn. Hoboken: John Wiley and Sons.
- Krosnick, Jon. 2000. "The Threat of Satisficing in Surveys: The Shortcuts Respondents Take in Answering Questions," *Survey Methods Newsletter* 20(1):4-8.
- Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 2005. "The Effects of Dependent Interviewing on Responses to Questions on Income Sources," *DIW Berlin Discussion Papers* 487.
- Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 2012. "The Impact of Questioning Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence From a Validation Study," *Journal of the Royal Statistical Society* 175(1):289-308.
- Moore, Jeffrey. 2006. "The Effects of Questionnaire Design Changes on General Income Amount Nonresponse in Waves 1 and 2 of the 2004 SIPP Panel." *U.S. Census Bureau Research Report Series: Survey Methodology* 2006-4.
- Rousseeuw, Peter, Christophe Croux. 1993. "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association* 88(424):1273-1283.
- Sala, Emanuela, S. C. Noah Uhrig, Peter Lynn. 2011. "'It is Time Computers Do Clever Things!': The Impact of Dependent Interviewing on Interviewer Burden," *Field Methods* 23(1):3-23.
- U.S. Census Bureau. 2016a. "Dependent Interviewing in the American Housing Survey." *American Housing Survey Help Guides and Other Documentation*. Available at https://www.census.gov/content/dam/Census/programs-surveys/ahs/tech-documentation/Dependent_Interviewing09.pdf.

- U.S. Census Bureau. 2016b. "Survey of Income and Program Participation 2014 Panel Users' Guide." *SIPP Users' Guide*. Available at <https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/2014-SIPP-Panel-Users-Guide.pdf>.
- Weinberg, Daniel. 2002. "The Survey of Income and Program Participation – Recent History and Future Developments," *SIPP Working Paper* No. 232.

Figure 1. Schematic of responses to a question with a $k=3$ design over six waves of a survey. The dependent interviewing version of the question, denoted by a red circle, is asked in the next three waves following a “yes” response.



Blue circle "Does your home have central air conditioning?"

Red circle "Last time we recorded that your home had central air conditioning. Is this information still correct?"

Figure 2. Percentage change from previous wave in estimated prevalence of housing characteristics, for all housing unit-level yes/no questions in the American Housing Survey occurring in at least two adjacent waves, 2009-2015 ($n_{2009}=122$, $n_{2011}=94$, $n_{2013}=149$, $n_{2015}=46$). Three questions had percentage change values greater than ± 80 percent in 2015 (-107, -543 and -1776 percent) and are not shown. Source: U.S. Census Bureau, American Housing Survey, 2007-2015. These data are subject to error arising from a variety of sources.

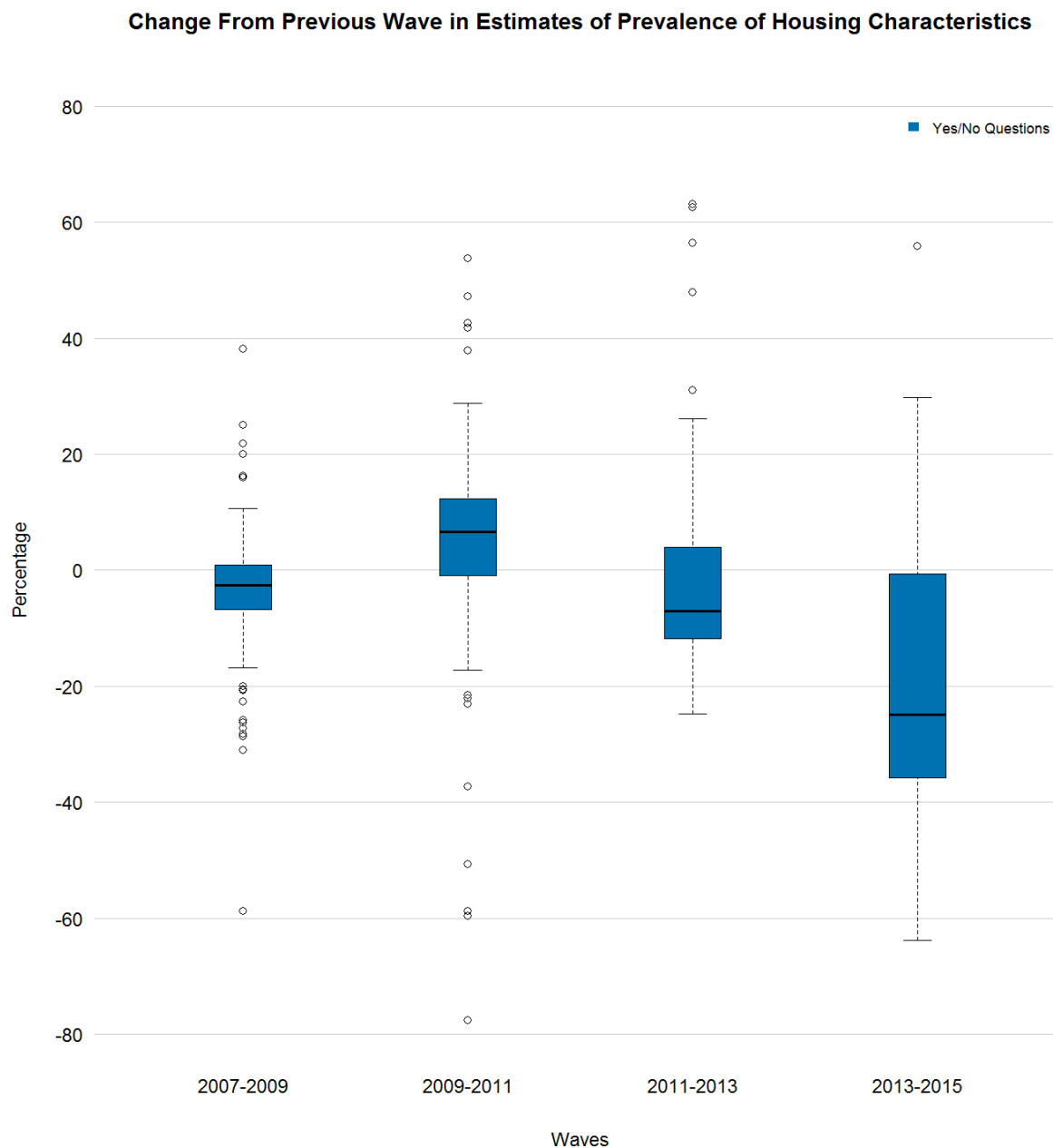
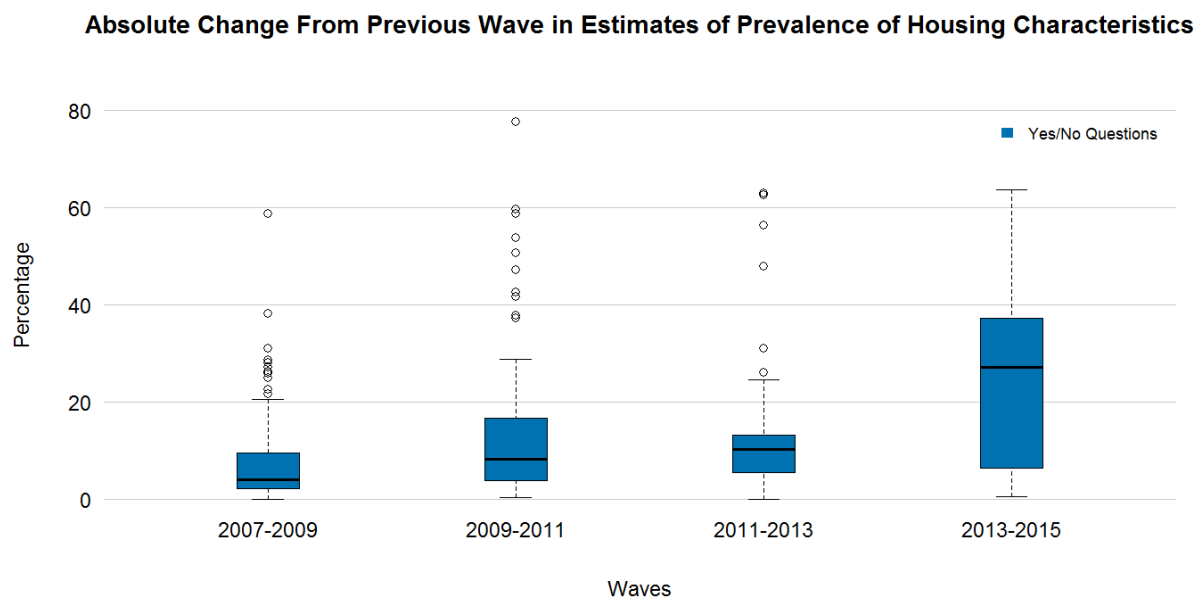
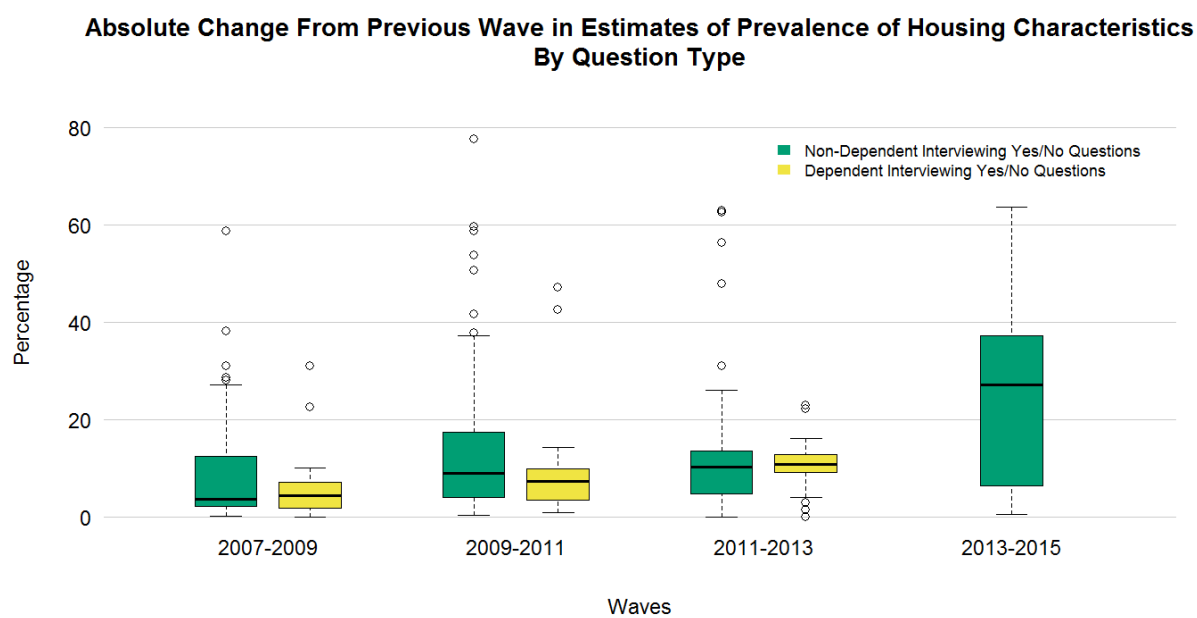


Figure 3. Absolute percentage change from previous wave in estimated prevalence of housing characteristics, (a) for all housing unit-level yes/no questions in the American Housing Survey, 2009-2015; (b) by whether question was a dependent interviewing question ($n_{2009,DI}=36$, $n_{2011,DI}=18$, $n_{2013,DI}=27$, $n_{2015,DI}=0$). Three questions had absolute percentage change values greater than 80 percent in 2015 (107, 543 and 1776 percent) and are not shown. Source: U.S. Census Bureau, American Housing Survey, 2007-2015. These data are subject to error arising from a variety of sources.



(a)



(b)

Figure 4. Percentage of housing units reporting a different decade built value or responding “I don’t know” in 2009 (n=32,416), for each 2007 decade built category. Percentages are unweighted. Source: U.S. Census Bureau, American Housing Survey, 2007-2009.

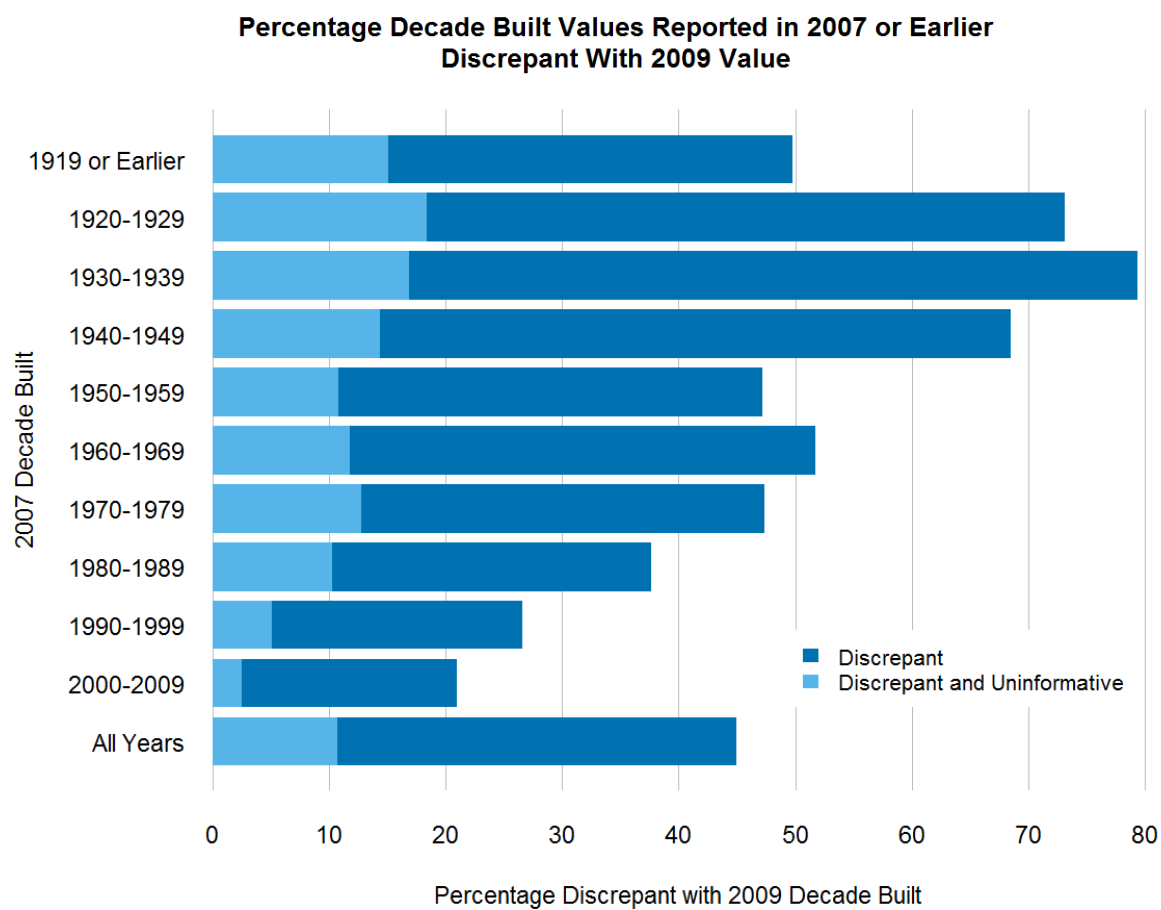


Figure 5. Decade structure was built, reported in 2007 and 2009 (n=32,416). Darker cells contain a larger proportion of cases. Proportions are row-wise, and unweighted. Source: U.S. Census Bureau, American Housing Survey, 2007-2015.

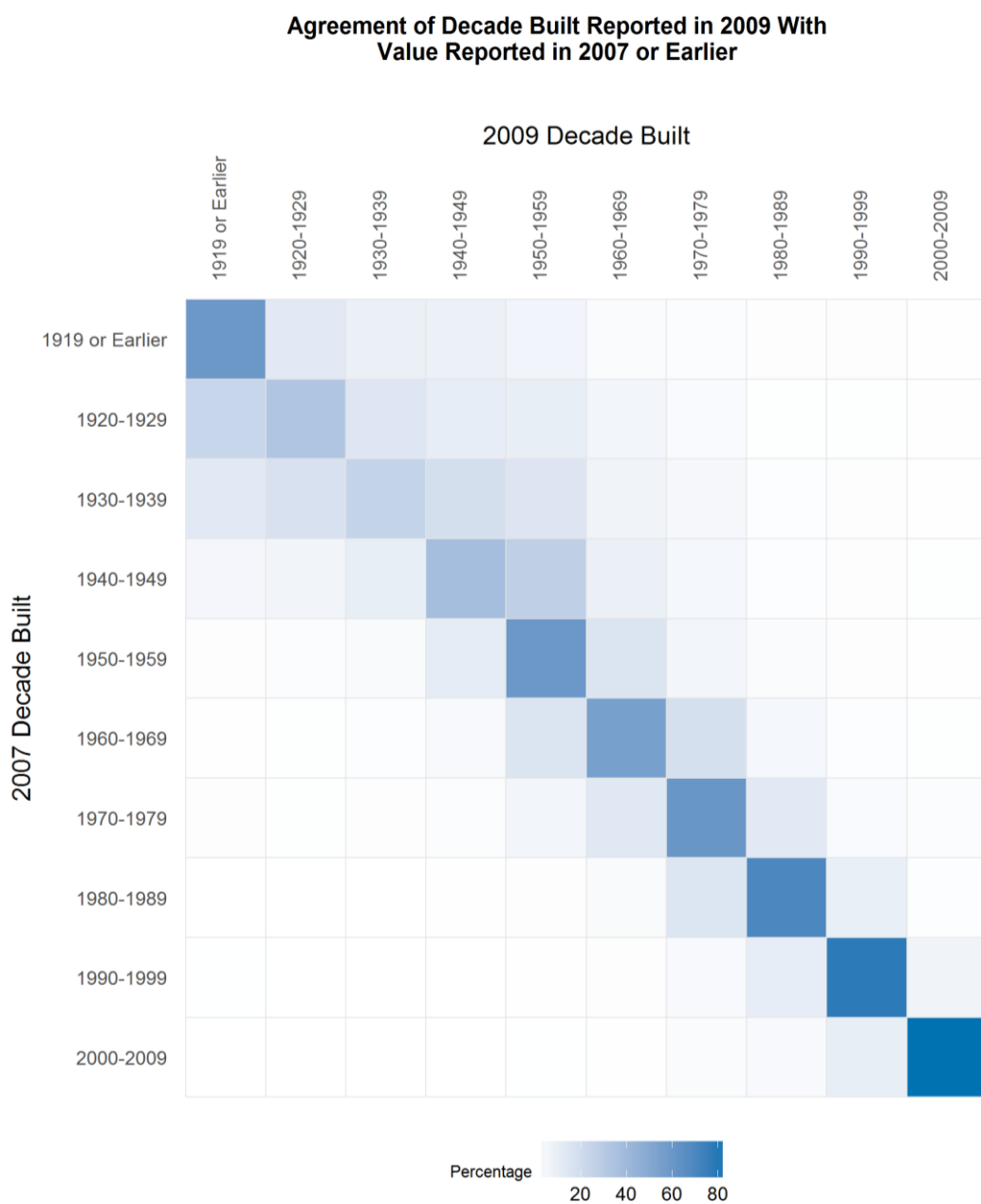


Figure 6. Adjusted Wald 95% confidence intervals for the difference in matched proportions of uninformative answers in 2013 and 2015, for all housing unit-level questions occurring in both waves and having at least 25 respondents answering in both waves ($n_{\text{questions}}=144$). Intervals above zero are from questions with a significantly increased percentage uninformative in 2015, and intervals below zero are from questions with a significantly decreased percentage in 2015. Responses are from the bridge sample ($n_{\text{respondents}}=3,908$). Source: U.S. Census Bureau, American Housing Survey, 2011-2015.

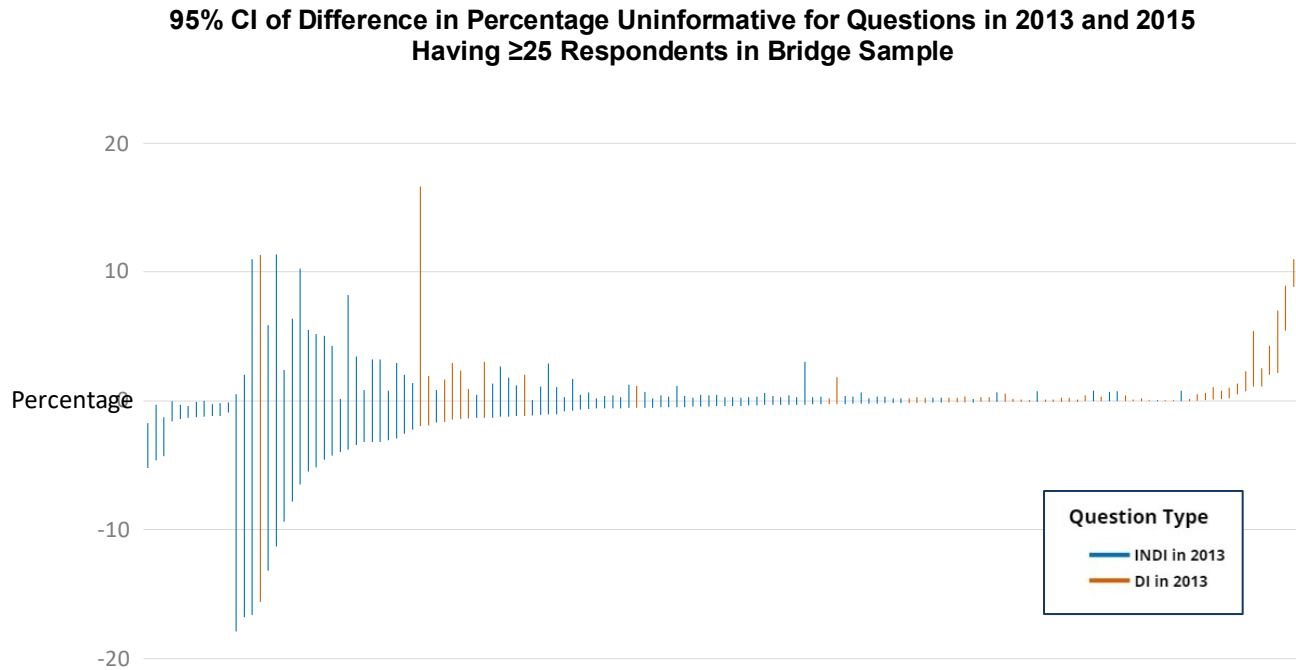


Figure 7. Percentage discrepant responses to questions on number of rooms in the unit of each of 12 types, 2009-2013, from housing units having a room count recorded in all three waves. Source: U.S. Census Bureau, American Housing Survey, 2009-2013.

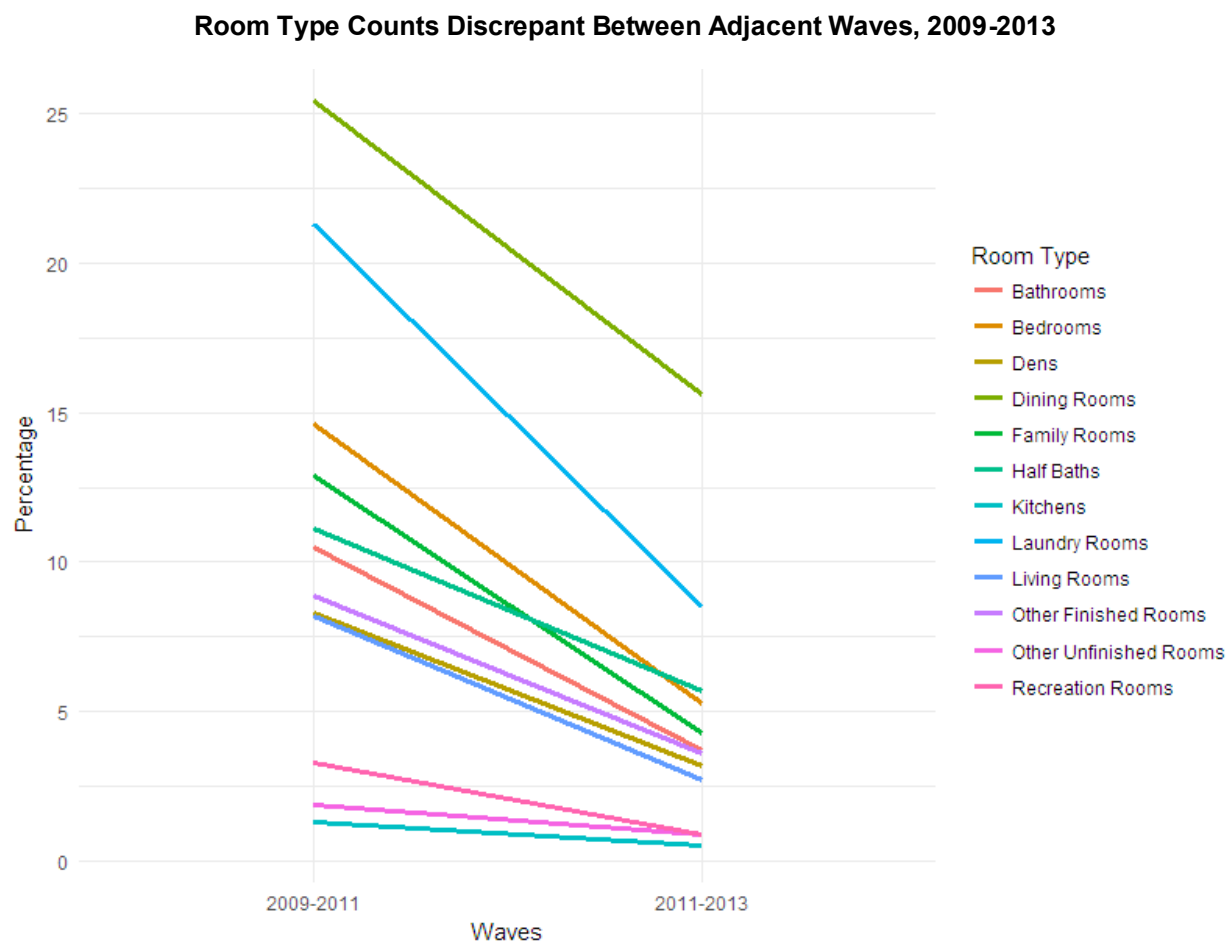


Figure 8. Flow diagram of change in responses to the question on number of dining rooms in the unit, 2007-2013, from housing units having a dining rooms count recorded in all four waves (n=24,242). Height of each bar represents the proportion of respondents reporting the given number of dining rooms. Source: U.S. Census Bureau, American Housing Survey, 2007-2013.

**Change in Proportion of Housing Units Reporting 0, 1, or 2 or More Dining Rooms,
2007-2013**

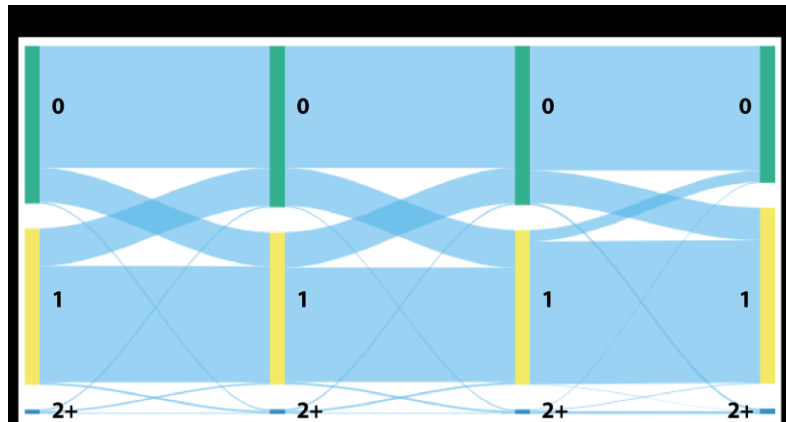


Table 1. Median and robust estimators of dispersion - median absolute deviation from the median (MAD) and S_n - for the absolute percentage change from previous wave in estimated prevalence of housing characteristics derived from yes/no questions, by whether questions were dependently interviewed (DI) or not (INDI) (Rousseeuw and Croux 1993). Source: U.S. Census Bureau, American Housing Survey, 2007-2015.

	2007-2009		2009-2011		2011-2013		2013-2015	
	<i>INDI</i>	<i>DI</i>	<i>INDI</i>	<i>DI</i>	<i>INDI</i>	<i>DI</i>	<i>INDI</i>	<i>DI</i>
Median	3.77	4.39	8.99	7.36	10.29	10.89	27.13	
MAD	2.36	2.66	6.24	3.35	4.59	1.89	14.34	
S_n	4.20	3.62	9.02	5.95	6.51	3.68	21.35	
n	86	36	76	18	122	27	46	0

Table 2. Categories used in 2007 AHS to record the year the structure was built, and the decade that these categories were collapsed to for analysis of consistency with 2009 responses. Source: U.S. Census Bureau, American Housing Survey, 2007.

2007 Year Built Category	Decade
1919 or earlier	1919 or earlier
1920-1929	1920-1929
1930-1939	1930-1939
1940-1949	1940-1949
1950-1959	1950-1959
1960-1969	1960-1969
1970-1974	1970-1979
1975-1979	
1980-1984	1980-1989
1985-1989	
1990-1994	1990-1999
1995-1999	
2000	2000-2009
2001	
2002	
2003	
2004	
2005	
2006	
2007	

Table 3. Percentage distribution of housing units with reported year built values discrepant between 2007 and 2009 (n=32,416). Percentages are unweighted. Source: U.S. Census Bureau, American Housing Survey, 2007-2009.

Year Built Values Reported in 2007 and 2009	Percentage
Year categories match	45.1
Year categories do not match	54.9
Different year category in 2009	43.4
“Don’t know” in 2009	10.7
Missing in 2009	0.8
Decades match	55.1
Decades do not match	45.0
Different decade in 2009	33.5
“Don’t know” in 2009	10.7
Missing in 2009	0.8

Table 4. Values reported in 2007 or earlier and in 2009 for decade structure was built (n=32,416). Percentages are unweighted. Source: U.S. Census Bureau, American Housing Survey, 2007-2009.

Decade Reported in 2007 or Earlier	Decade Reported in 2009	
	<i>Most common (%)</i>	<i>Percentage within one decade of 2007 value</i>
1919 or earlier	1919 or earlier (50.2)	60.4
1920-1929	1920-1929 (26.9)	57.1
1930-1939	1930-1939 (20.6)	49.3
1940-1949	1940-1949 (31.5)	62.4
1950-1959	1950-1959 (52.8)	75.7
1960-1969	1960-1969 (48.3)	77.0
1970-1979	1970-1979 (52.7)	74.3
1980-1989	1980-1989 (62.3)	83.5
1990-1999	1990-1999 (73.4)	88.4
2000-2007	2000-2009 (79.9)	88.1

Table 5. Distribution of housing unit-level questions asked of a subsample (n=3,908) of respondents in 2013 and 2015, by whether the question was a dependent interviewing (DI) question in 2013 and whether its percentage uninformative responses in 2015 were significantly different from the percentage uninformative in 2013 (95% adjusted Wald CI for difference in matched pairs proportions did not contain zero). Only questions having at least 25 respondents answering in both waves were included in the analysis. Source: U.S. Census Bureau, American Housing Survey, 2013-2015.

2013 Question Type	2015 Percentage Uninformative			Total
	<i>Decreased</i>	<i>No Change</i>	<i>Increased</i>	
INDI	11	82	0	93
DI	0	39	12	51
Total	11	121	12	144

Table 6. Percentage of responses to room count questions that were discrepant with the count reported in the previous wave, for 2011 and 2013. Differences in percentages for each room type were significant ($p < 0.0001$). Respondents were asked independent (INDI) questions in 2009 and 2011, and dependent interviewing (DI) questions in 2013 for counts of each room type. Cases with a missing count for a room type in any wave were excluded from that room type's percentage calculations. Percentages are unweighted. Source: U.S. Census Bureau, American Housing Survey, 2009-2013.

Room Type	n	Percentage Responses Discrepant	
		2009-2011 (INDI)	2011-2013 (DI)
Dining Rooms	27,969	25.4	15.6
Laundry Rooms	18,295	21.3	8.5
Bedrooms	27,988	14.6	5.3
Family Rooms	14,904	12.9	4.3
Half Baths	27,948	11.1	5.7
Bathrooms	27,997	10.5	3.7
Other Finished Rooms	11,823	8.9	3.6
Dens	12,571	8.3	3.2
Living Rooms	28,019	8.2	2.7
Recreation Rooms	10,773	3.3	0.9
Other Unfinished Rooms	10,418	1.9	0.9
Kitchens	28,047	1.3	0.5

Table 7. Change in percentage of national sample housing units reporting zero or one dining room. Only housing units having both a completed interview and a numeric answer for number of dining rooms in all waves 2007-2013 were included (n=23,926). Percentages are unweighted. Source: U.S. Census Bureau, American Housing Survey, 2013-2015.

Reported Number of Dining Rooms		Percentage		
<i>Wave 1</i>	<i>Wave 2</i>	<i>2007-2009</i>	<i>2009-2011</i>	<i>2011-2013</i>
0	0	38.5	38.5	38.2
0	1	10.7	11.8	10.2
1	0	11.8	11.1	3.7
1	1	36.6	36.1	44.8

Table 8. Odds ratio estimates for explanatory variables in a model of whether an answer to a fuel question was discrepant with the previous wave answer. Source: U.S. Census Bureau, American Housing Survey, 2011-2015.

Response Variable	Explanatory Variables	Odds Ratio Estimate (95% CI)	Fixed Effect Significance Test P-Value	
Answer was discrepant with answer from previous wave	Tenure		<0.0001	
	Renter	.		
	Owner	0.6 (0.49,0.75)		
	Structure type		0.0167	
	Single unit detached	.		
	Single unit attached	1.05 (0.76, 1.46)		
	Multiunit building	1.23 (0.98, 1.54)		
	Manufactured/mobile home	0.59 (0.38, 0.91)		
	Fuel question asked		<0.0001	
	Main cooking fuel	.		
	Clothes dryer fuel	1.46 (1.16, 1.85)		
	Main heating fuel	3.05 (2.52, 3.69)		
	Water heater fuel	2.72 (2.42, 3.30)		
	Respondent was asked the independent question		41.3 (25.5, 66.8)	<0.0001