

August 5, 2020

---

American Housing Survey

# Structure Type Estimates: 2017 and 2019

---

Version 1.0

DRB Clearance Number - CBDRB-FY20-POP001-0175

DRB Clearance Number - CBDRB-FY20-POP001-0175

## 1. Introduction

Initial American Housing Survey (AHS) estimates in 2017 and 2019 showed a sharp increase in single attached units compared to 2015, and a decrease in housing units in multi-unit structures having fewer than 20 apartments (Table 1), while American Community Survey (ACS) estimates in the same time period were stable. After ruling out a number of possible causes, Census concluded that the increases may be due to questionnaire changes coupled with respondents' difficulty distinguishing single attached units from multi-units.

Because structure type is fundamental to AHS data processing and presentation, and because the observed increases were likely due to error, structure type responses were edited in the 2017 and 2019 data for units reporting as single attached but having characteristics of a multi-unit, and vice versa. These changes were made based either on a set of criteria or on the output of a model that assigned structure type. This brought the national estimates and most metro-level estimates of single attached units and of multi-units closer to ACS estimates.

This document summarizes the 2015-2019 structure type estimates and the investigation of potential causes of the observed changes. It details the steps taken to stabilize the 2017 and 2019 structure type estimates – prior to releasing the 2017 estimates, in creating the 2019 instrument, and prior to releasing the 2019 estimates – and outlines the plan for the 2021 instrument.

**Table 1.** National structure type estimates 2013-2019; 2017 and 2019 estimates are prior to editing. Estimates are in thousands. Due to rounding, totals may not equal the sum of their components.

	AHS			
	2013	2015	2017	2019
Total	132,800	134,800	137,400	139,700
Single Attached Unit	7,560	9,840	12,550	11,440
Multi-unit	33,030	33,050	31,550	32,500
2 Units	4,898	4,664	3,788	3,909
3-4 Units	5,740	5,698	4,924	5,162
5-9 Units	6,564	6,075	5,713	5,917
10-19 Units	6,109	5,555	5,342	5,380
20-49 Units	4,558	4,727	4,506	5,028
50+ Units	5,163	6,327	7,275	7,111
Single Detached Unit	83,310	83,150	84,820	87,370
Mobile Home	8,603	8,686	8,407	8,272
Other	33	69	75	96

## 2. History of Changes to AHS Structure Type Questions, 1983-2019

The AHS defines a single attached unit as a housing unit that has unbroken walls extending from the ground to the roof that divide it from adjoining structures. If pipes or duct work pass from one unit to

another then the walls would be broken and the AHS would categorize it as a multi-unit structure. In 1984, research on AHS determined that the 1973-1983 surveys were consistently overestimating the number of single attached units (see the Introduction to the AHS 1985 summary tables book<sup>1</sup>, *Qualifications of the Data* section, *Units in Structure* subsection). Census found that respondents in this period had difficulty distinguishing between single attached units and multi-unit structures. Figure 1 below shows how the question was phrased in 1983.

Figure 1. AHS National 1983 questionnaire, structure type question.

8a. How many living quarters, both occupied and vacant, are there in this house (building)?

101

1  Mobile home or trailer (no permanent room attached) - Go to 8b

2  One, detached from any other building (includes mobile home or trailer WITH one or more permanent rooms attached)

3  One, attached to one or more buildings

4  2

5  3 or 4

6  5 to 9

7  10 to 19

8  20 to 49

9  50 or more

Skip to 8c

Skip to 9a

OFFICE USE ONLY

101

x

Questions for single family units that could lead to re-categorizing the structure type as multi-units were added to the AHS 1984 metropolitan instrument and 1985 national instrument (Figure 2). Though there were many questions impacting structure type in 1985, over the course of the panel they were gradually removed or made independent of structure type assignments.

By 2013, only two such questions were asked of single attached units: 1) if the unit shared attic or basement space with another unit (SHARAT), and 2) if there was any shared furnace or boiler equipment (SHARFR). An affirmative answer to either of these items would indicate that the unit did not have an unbroken wall separating it from the adjoining unit, and the unit would be edited to a multi-unit structure type.

The AHS was redesigned in 2015. This involved drawing a new sample to replace the 1985-2013 panel, and updating the questionnaire. As part of a broader effort to reduce respondent burden, the shared furnace question SHARFR was removed in 2015 and the shared attic question SHARAT was removed in 2017.

In 2019, changes were made to the instrument in order to investigate and address respondents' misidentification of multi-units and single attached units. SHARAT and SHARFR were added back in, and

<sup>1</sup> <https://www2.census.gov/prod2/ahsscans/h150-85.pdf>, page X.

two other questions were added that did not edit structure type responses in the instrument, but did provide additional information to use in structure type editing. ABUNIT1 and ABUNIT2 asked renters in single attached units and multi-units, respectively, whether there was another housing unit above or below theirs. While any single attached unit reporting another unit above or below could be categorized as a multi-unit, a multi-unit reporting no unit above or below could not be assumed to be single attached, because it was not known whether the unit had unbroken walls from ground to roof.

Figure 2. National AHS 1985 questionnaire, structure type questions.

<b>REGULAR OCCUPIED</b>	
<p><b>20.</b> Are your living quarters in a -- <i>(Read answer categories.)</i></p>	<p style="text-align: center;"><b>1120</b></p> <p>1 <input type="checkbox"/> Mobile home?                  2 <input type="checkbox"/> One-unit building detached from any other building?                  3 <input type="checkbox"/> One-unit building attached to one or more buildings? — Skip to item 22a                  4 <input type="checkbox"/> Building with two or more apartments? — Skip to item 21b</p>
<p><b>21a.</b> Are there any occupied or vacant apartments besides your own in the (building/mobile home)?</p>	<p style="text-align: center;"><b>1130</b></p> <p>1 <input type="checkbox"/> Yes — Fill Table X on Control Card, then go to item 21b                  2 <input type="checkbox"/> No — Skip to item 23 and mark box 1 or 4</p>
<p><b>b.</b> How many apartments are in the (building/mobile home)?</p>	<p style="text-align: center;"><b>1140</b></p> <p>_____ Number — Skip to item 23 and mark box 3 or 5</p>
<p><b>22a.</b> Does your (house/apartment) share an attic or basement with the (house/apartment) next door?</p>	<p style="text-align: center;"><b>1150</b></p> <p>1 <input type="checkbox"/> Yes                  2 <input type="checkbox"/> No . . . . .                  3 <input type="checkbox"/> Don't know } SKIP to item 22c</p>
<p><b>b.</b> How many (houses/apartments) share the attic or basement?</p>	<p style="text-align: center;"><b>1160</b></p> <p>_____ Number. — If one, reask item 22a and correct entry.                  If more than one, skip to item 23 and mark box 3.</p>
<p><b>c.</b> Does your (house/apartment) share a furnace or boiler with the (house/apartment) next door?</p>	<p style="text-align: center;"><b>1170</b></p> <p>1 <input type="checkbox"/> Yes                  2 <input type="checkbox"/> No . . . . .                  3 <input type="checkbox"/> Don't know } SKIP to item 22e</p>
<p><b>d.</b> How many (houses/apartments) share the furnace or boiler?</p>	<p style="text-align: center;"><b>1180</b></p> <p>_____ Number — If one, reask item 22c and correct entry.                  If more than one, skip to item 23 and mark box 3.</p>
<p><b>e.</b> Are there any occupied or vacant apartments besides your own in the building?</p>	<p style="text-align: center;"><b>1190</b></p> <p>1 <input type="checkbox"/> Yes — Fill Table X on Control Card, then go to item 22f                  2 <input type="checkbox"/> No — Skip to item 23 and mark box 2</p>
<p><b>f.</b> How many apartments are in the building?</p>	<p style="text-align: center;"><b>1200</b></p> <p>_____ Number — If one, reask item 22e and correct entry.                  If more than one, go to item 23 and mark box 3.</p>
<p><b>23.</b> Number of units in building based on entries in items 20–22.</p>	<p style="text-align: center;"><b>1210</b></p> <p>1 <input type="checkbox"/> One-unit building—detached                  2 <input type="checkbox"/> One-unit building—attached                  3 <input type="checkbox"/> Two-or-more-unit building . . . . .                  4 <input type="checkbox"/> One-unit mobile home . . . . .                  5 <input type="checkbox"/> Two-or-more-unit mobile home } Skip to item 25a</p>

### 3. Possible Sources of Error in Structure Type Estimates

Census considered the following possible sources of error in the 2015-2019 data, ruled out most of them, and found inconclusive evidence for the rest.

## Instrument Error

To determine whether an instrument error may have changed the structure type during some interviews in 2019, Census compared the keyed structure type responses in the instrument keystroke log files to the structure type in the datasets of responses produced from other instrument output data files, which are processed to become the IUF. Fewer than 75 respondents' records had a discrepancy, so the instrument was excluded from further investigation.

## Backtracking in the Instrument

After the main structure type question ISTYPE is asked ("Are your living quarters in a: Manufactured/Mobile home? One-unit building, detached from any other building? One unit building, attached to one or more buildings? Building with two or more apartments?"), subsequent questions may indicate to the interviewer that the wrong structure type was entered at ISTYPE. Some questions are asked of only one structure type (for example, COMPLEX, asked of multi-units: "Some apartment buildings are part of a complex of multiple apartment buildings. Is this building part of a complex?"), or error screens may alert the interviewer to a discrepancy between the current question's response and the previously reported ISTYPE. These could have caused interviewers to go back to ISTYPE and change the answer.

Census used the instrument keystroke files to identify when the interviewer backtracked this way. The percentage of interviews where ISTYPE was changed at least once was between 7.3% and 8.5% in 2015, 2017 and 2019. In each wave, fewer than 1000 interviewed units switched between the single attached and multi-unit categories. Of these, the net movement into one category in each year was around 100 units. In 2015 and 2017, the single attached unit category netted about 100 units from backtracking. In 2019, the multi-unit category netted about 100 units. Because each wave's net change was rather small, backtracking was excluded as a possible source of error.

## Data Processing Error

Each survey response goes through multiple steps of processing, which includes editing item responses for consistency with other items, creating new variables, and imputing missing values for some variables. This introduces the possibility of an error in data processing causing incorrect values to be recorded for a variable. We tracked each respondent's structure type value across the processing steps in the 2019 data, and found that processing changed fewer than 100 structure type responses, so this was ruled out as a possible cause in 2019.

## Removal of SHARFR in 2015

Census considered the possibility that the removal of the shared furnace question (SHARFR) in 2015 could explain some of the shift from 2013, but early research was inconclusive. In practice, it was difficult to collect enough evidence for three reasons:

1. Since 2001, returning households weren't asked structure type again if they gave a valid response in the prior year, and in the two years before that they were asked as verification of prior responses. This made it difficult to isolate the impact of SHARFR because households could be carrying forward prior responses that were influenced by versions of the questionnaire that were closer to Figure 2.

2. Not all responses to SHARFR were output into the raw data. In many instances, when a response to SHARFR triggered an error screen or an edit, the value was cleared out if the respondent changed their structure type. This problem can be resolved by pulling from interviewer keystroke paradata, but adds some complexity to the problem.
3. The size of the bridge sample - 6,000 respondents from the 1985-2013 panel surveyed in 2015 - was a limitation. While it was sufficient for studying the impact of weighting and sample design on broad AHS estimates, the number of units shifting structure types between these two categories was small. The bridge sample also fully used the 2015 questionnaire due to the complexities of managing multiple collection instruments in the field. Ideally, to study the impact of question changes it would have been split to have both versions of the questions represented in 2015.

Reasons 1 and 3 are not a factor when analyzing the impact of removing SHARAT on the 2017 data because they each apply to the full sample and there was no dependent interviewing for the structure type variables in 2017.

Census examined the 2015 keystroke paradata and identified every time an interviewer changed the structure type from single attached unit to another structure type after hitting the SHARAT screen. Table 2 shows the impact on national estimates of forcing all of these responses back to single attached units.

**Table 2.** Effect on AHS 2015 national estimates of reversing 2015 structure type reporting changes that occurred in the instrument after the interviewer viewed SHARAT (in thousands).

	<b>AHS 2015</b>	
	<i>Total</i>	<i>Adjusted</i>
Single Attached Unit	9,840	11,100
Multi-unit	33,050	32,080
Single Detached Unit	83,150	82,920
Mobile Home	8,686	8,683

Though Table 2 shows significant movement between structure types during the interview, it's not clear how much of this results directly from the SHARAT question itself. It is possible to identify many of the shifts with direct edits from SHARAT responses, but there are other patterns that are more difficult to quantify. For example, an interviewer might backtrack once entering the SHARAT screen without ever answering it. They could also answer it in a way that triggers an error screen rather than a direct edit, or backtrack once they realize a later line of questioning doesn't make sense. It is possible that the table above overstates the specific impact of SHARAT because some respondents may have changed responses even if the question was not there.

In addition to quantifying the impact of the question, Census also wanted to determine whether the SHARAT edit was improving structure type categorization even when it conflicted with the respondent's initial answer. To better understand this, a sample of units that were directly edited by their responses to SHARAT in 2015 was drawn. Two Census analysts independently made structure type assignments on



a range of qualitative data, including interviewer notes, write-in responses, address fields, physical descriptions, and satellite imagery of the area. Of those with enough information to make an assignment, 70.8% were identified as multi-units and 29.2% were identified as single attached or single detached. Though there was some ambiguity with these assignments, Census believes that the inclusion of this question on the whole is more beneficial than not.

In an analysis of the 2019 instrument keystroke files, it was found that together SHARAT and SHARFR reduced the national estimate of single attached structures by 684,900, by changing about 350 units' structure type, with SHARAT accounting for 85% of those.

#### Removal of Dependent Interviewing in 2015

ISTYPE was a "Do Not Re-ask" dependent interviewing question through 2013, meaning that respondents in units having a response recorded in the previous wave were not asked the question again in the next wave. The exception was that if a single attached unit's response to SHARAT or SHARFR indicated it was a multi-unit, then ISTYPE was re-asked. Dependent interviewing was removed for ISTYPE starting in 2015 with the new sample.

The 2015 bridge sample of 6,000 units from the 1985-2013 panel had a high level of movement between single attached unit and small multi-unit structure types, relative to other categories, from 2013 to 2015. It displayed an increase of 39% in estimated single attached units between 2013 and 2015 (around 250 bridge sample units changed from multi-unit or single detached in 2013 to single attached in 2015). This indicates that dependent interviewing through 2013 may have provided stability to a question that for at least some answer categories would not otherwise have a stable distribution across waves. The absence of dependent interviewing for the structure type question in 2015, 2017 and 2019 could have contributed to the instability of the structure type distribution in those waves.

#### Respondent Understanding of Structure Type Categories

For the approximately 30,000 housing units having the same household reporting in 2015, 2017 and 2019, 84% reported the same structure type in all three waves. Of those reporting a structure type of single attached at least once, less than 25% reported single attached in all three waves, indicating a possible lack of understanding of what defines a single attached unit. This is a plausible source of error, since single attached unit is a real estate concept, and is not defined for the respondent during the interview. A definition of single attached unit can be found the instrument's help screen for ISTYPE, yet help screens are used rarely in the AHS. Moreover, help screens might not be used in situations where a respondent provides an answer without understanding the question.

Removal of dependent interviewing, along with the difficulty of understanding the difference between "One-unit building, attached to one or more buildings" and "Building with two or more apartments" and a tendency of respondents in multi-units to choose the former, might together account for much of the error in the estimated percentage of single attached units.

## 4. 2017 AHS: Edits to Structure Type Responses

For the 2017 AHS wave, Census implemented two types of edits to the structure type variable. In the national sample and 15 largest metropolitan areas, we implemented rule based edits using data from the 2015 and 2017 waves. In the 10 rotating metropolitan areas that rotated into the sample in 2017 and did not have 2015 data, we implemented a model-based edit. The sections below describe each of these edits in more detail.

### National and 15 Largest Metropolitan Areas Data: Rule-Based Edit

A sample of units shifting between single attached and multi-unit responses from 2015 to 2017 was drawn, to which Census assigned a structure type based on qualitative variables. One of the strongest indicators of structure type for this group was the presence of an apartment number in the address. Over 75% of those in this group with apartment numbers were assigned as multi-units in the qualitative checks.

Using this information from the qualitative review, Census created an edit that would determine the structure type assignments for housing units that gave conflicting responses between single attached unit and multi-unit in the 2015 and 2017 surveys. Table 3 shows how the edit determined which structure type to assign and which variables were carried forward from the prior year when the 2017 value was edited.

The edit was applied only to housing units where responses shifted between single attached unit and multi-unit. Because we rarely expect structure type to change without extensive reconstruction of the unit, the edit effectively decided which of two conflicting responses to use rather than editing anything that was consistently reported between the 2015 and 2017 waves. After the edit, all variables directly related to structure type not included in Table 3 were edited around the new structure type as if it were reported that way.

**Table 3.** 2017 structure type assignments for responses from the national and largest 15 metropolitan areas' samples that conflicted with 2015.

<b>Edit of Structure Type</b>	<b>Assignment Criteria</b>	<b>Variables Carried Forward from 2015</b>
From multi-unit to single attached unit	Does not have an apartment number and number of stories is 1, 2, or 3.	BLD, FOUNDTYPE, LOT, SUBDIV
From single attached unit to multi-unit	Has an apartment number or number of stories is greater than 3.	BLD, COMPLEX, KITEXCLU, MGRONSITE, NUNITS, STORIES

Table 4 below shows the weighted results of the adjustments along with the 2017 ACS. We see that the edit closed much of the gap between the 2015 and 2017 estimates.

**Table 4.** Effect of edit on 2017 national structure type estimates (in thousands).

	AHS 2015	AHS 2017		ACS 2017
		Original	Edited	
Single Attached Unit	9,840	12,550	10,040	8,057
Multi-unit	33,050	31,550	34,060	35,990
2 Units	4,664	3,788	4,060	4,915
3-4 Units	5,698	4,924	5,684	5,991
5-9 Units	6,075	5,713	6,645	6,425
10-19 Units	5,555	5,342	6,180	6,066
20-49 Units	4,727	4,506	4,903	
50+ Units	6,327	7,275	6,593	
20+ Units	11,050	11,780	11,500	12,590
Single Detached Unit	83,150	84,820	84,830	84,730
Mobile Home	8,686	8,407	8,397	8,500
Other	69	75	75	128

### Rotating Metropolitan Areas Data: Model-Based Edit

The 2017 AHS sample included 10 metro areas not sampled in 2015, referred to as the rotating metros. Since the edit to the national and largest 15 metro area data used 2015 responses, it could not be applied to the rotating metros' data. As with the national estimates, the rotating metros exhibited a gap between the initial 2017 AHS estimates and the 2017 ACS estimates of single attached units and multi-units (Table 5).

Census assessed several classification model options to help re-assign structure type in the 2017 rotating metro data using knowledge gained from the national data's editing process. The training data for these classification models was composed of units in the 2017 national sample listed as single attached or multi-unit and which were in both the 2015 and 2017 data. The training data included a wide range of features based on questions asked of both single attached unit and multi-unit groups. For cross validation, the training data were split into training (70%) and testing (30%) subsets. The prediction dataset was composed of the 2017 metro sample units listed as single attached unit or multi-unit, with the same features as the training data.

The objective of using a classification model was to re-assign units in the metro data that were similar to the units re-assigned in the national data. Census trained and assessed several classification models. Assessment of the models included:

- How well does the model predict the withheld test in the training dataset?
- How does the proportion of units reassigned in the metro data align with the proportion of units reassigned in the national data?

- Do weighted estimates at the metro level align with ACS 2016 estimates<sup>2</sup>?

**Table 5.** AHS 2017 structure type estimates (in thousands) for the rotating metros before and after a model based edit, compared to American Community Survey 2017. SFA=single attached unit; MU=multi-unit.

	Structure Type	AHS 2017		ACS 2017
		Original	Edited	
Baltimore-Columbia-Towson, MD	SFA	414.5	313.1	342.0
	MU	213.4	314.9	287.0
Birmingham-Hoover, AL	SFA	43.6	22.5	11.2
	MU	65.8	86.8	99.8
Las Vegas-Henderson-Paradise, NV	SFA	116.8	39.4	42.1
	MU	232.1	309.5	294.8
Minneapolis-St. Paul-Bloomington, MN-WI	SFA	155.3	148.1	148.0
	MU	362.2	369.4	384.0
Oklahoma City, OK	SFA	38.1	13.0	14.5
	MU	91.9	117.0	115.8
Richmond, VA	SFA	65.1	30.6	37.4
	MU	92.9	127.5	107.2
Rochester, NY	SFA	42.3	36.6	24.9
	MU	114.6	120.3	126.1
San Antonio-New Braunfels, TX	SFA	56.6	18.9	19.1
	MU	190.9	228.9	219.4
San Jose-Sunnyvale-Santa Clara, CA	SFA	95.9	47.6	72.0
	MU	204.9	253.2	228.3
Tampa-St. Petersburg-Clearwater, FL	SFA	154.8	101.8	87.8
	MU	330.1	383.1	390.6

Models were optimized using a random grid search algorithm to find model parameters that led to the best score, defined as the highest f1-weighted score. The minority outcome units were oversampled (sampling with replacement) during the model optimization and training process. This was done to protect against model bias when predicting rare events.

The trained models were used to predict structure type for 2017 rotating metro sample units that were reported as either single attached units or units in multi-unit structures. Based on this, the best classifier was an Ada boosted decision tree with a 4-class outcome: 1) remained single attached unit, 2) remained multi-unit, 3) re-assigned from single attached unit to multi-unit, 4) re-assigned from multi-unit to single attached unit.

Using the predictions, an adjusted structure type was created using the rules in Table 6. These adjustments were designed based on cross-validation tests conducted on the withheld test data during

<sup>2</sup> 2017 ACS estimates were not available at the time of analysis.

the training process. The adjustment assumes the model correctly adjusted structure type but not the type of change.

The model assessment ensured that those reclassified in the metro dataset were similar to those in the national dataset. What this assessment *did not* do is assess whether those that were re-classified or kept the same were actually correct. To do such an assessment, we conducted a qualitative assessment of the re-classification process.

**Table 6.** AHS 2017 rotating metros structure type adjustment based on model predictions.

<b>Outcome</b>	<b>Prediction</b>	<b>Adjustment</b>
4-Class Outcome	Remained single attached unit	Single attached unit
	Remained multi-unit	Multi-unit
	Single attached unit to multi-unit	Multi-unit
	Multi-unit to single attached unit	Single attached unit

Census manually confirmed structure type for a random set of national and metro units. The objective of this validation was to set a benchmark to assess the metro re-classification process. The results demonstrated the error rates from the classification model aligned with the error rates from the national data reclassification process.

Table 4 shows the weighted results of the adjustments in the “AHS 2017 Edited” column, along with the ACS 2017. The edit closed much of the gap between the AHS and ACS estimates.

Re-assigning the metro structure type data occurred before the general AHS edits. This allowed for the standard edits to impute missing data or to mark data items for which a unit was no longer in universe based on its structure type as “Not Applicable.”

### Limitations of Approach

While the above edit was sufficient in the short term for the 2017 AHS data release, there were known limitations to this methodology. First, the national and top 15 metro data edit only applied to housing units that were asked questions related to structure type in both 2015 and 2017. This left out new construction and rotating metropolitan sample added in 2017.

Secondly, there were assumptions in this edit that were made based on qualitative observations of a small number of units. These decisions were made based on time limitations.

Thirdly, the metro data edit focused on re-classifying metro units that were most similar to units that were assigned in the national data. This excluded metro specific features that might have influenced the structure type selected.

These limitations led to the decision to add SHARAT, SHARFR, ABUNIT1 and ABUNIT2 to the 2019 instrument, discussed earlier, and to consideration of incorporating more third party administrative data into data processing in 2019 and beyond.

## 5. 2019 AHS: Edits to Structure Type Responses

Census sought a small set of AHS and administrative variables that could be used to edit structure type as in 2017. We found that a number of variables correlated with single attached unit or multi-unit structure types, but could not be used as inputs to direct edits of structure type because the correlation was not strong enough to provide satisfactory results. A logistic regression model was selected as the mechanism to assign a structure type of either single attached unit or multi-unit. Units selected for possible reassignment of structure type were those that had reported being a single attached unit or a multi-unit in 2019, and were either first interviewed in 2019 or had been interviewed in a previous wave and reported a 2019 structure type inconsistent with the unit's previous wave responses.

### Candidate Models Built and Evaluated

The dataset the candidate models were built on was composed of units Census considered to be “true multi-units” and “true single attached units,” having all of these characteristics (n=13,280):

- Responded to the survey in:
  - 2015, 2017 and 2019, or
  - 2015 and 2017 only, or
  - 2015 and 2019 only
- Reported the same structure type in all interviews
- Reported structure type of single attached unit or multi-unit only
- The 2019 Census master address file count of units at the same basic street address did not conflict with the reported structure type.

This dataset was partitioned into a 70/30 training/test split of randomly selected units. The binary response variable was whether the unit was a multi-unit or a single attached unit. The 17 prospective input variables came from unedited responses to other questions in the AHS and variables from third party administrative data, all categorical. There was item-level missingness both in the data used to train and select the model, and in the set of target units whose structure type was recategorized. Missing values were not imputed; item-level missing values and, within the administrative variables, unit-level missingness, each had their own response category.<sup>3</sup>

Model selection was done using forward selection on logistic regression models having no interaction terms and no transformations of input variables, with lowest AIC as the selection criterion at each step. Forward selection was stopped at 12 variables because model performance did not improve with the addition of any remaining variables.

### Model Selected

All models with 10 to 12 input variables had a multi-unit misclassification rate of around 1%, and a single attached misclassification rate of 31-33%. The 10-variable model with the lowest single attached misclassification rate among the 10-variable models was chosen for the sake of a simpler model.

---

<sup>3</sup> Using only complete cases for building the model would have resulted in a much smaller training set (about 1,500 units), and a possibly biased model that could not assign a structure type to units in the target set having missing input variable values.

The model selected had the following input variables:

DRB Clearance Number - CBDRB-FY20-POP001-0175

*From AHS*

- Another unit, or a space shared with another unit, is above or below this one<sup>4</sup>
- Unit is part of a home association<sup>5</sup>
- Unit has 2 or fewer floors
- Unit has a garage
- Unit street address has an apartment number
- Tenure of household<sup>6</sup>

*From third party administrative data, 2019 vintage<sup>7</sup>*

- Administrative data's structure type code, mapped to an AHS structure type
- Whether administrative data number of stories is greater than AHS number of floors
- Number of baths is 0 or missing
- Number of bedrooms is 0 or missing

## Results

The units whose structure type was subject to change by the model were those that answered the structure type question ISTYPE with "single attached" or "multi-unit" in 2019, and either had no prior wave response, or had a prior wave response, reported a different structure type from 2019, and that different structure type was either "single attached" or "multi-unit."

The model had an overall 5.4% misclassification rate. Its misclassification rate for multi-units was 1.1%, and for single attached units was 31.6%. All models tested tended to classify a large portion of single attached units as multi-units: the lowest single attached misclassification rate was above 30%.

Of the approximately 10,000 units from both the national and metro files that went through the model, 3,800 were recategorized. Of the 4,100 single attached units entering the model, 83% left the model as multi-units. For the multi-units entering the model, only 6% left as single attached.

In the national sample, 43% of single attached units were switched by the model to multi-units, and less than 2% of multi-units were switched to single attached.

Both the national estimates of single attached and multi-units and the majority of the metro estimates were closer to the ACS 2018 estimates with the model-based changes to structure type than without.

---

<sup>4</sup> Units in the training or test dataset that had responses in 2015 and 2017 only did not have a value. In 2019, only renter households whose primary residence was the sample unit were asked whether there was another unit above or below theirs.

<sup>5</sup> In 2019, only respondents who were homeowners and whose primary residence was the sample unit were asked about home association membership.

<sup>6</sup> Only occupied interviews – the unit was not vacant, and the occupants did not have a usual residence elsewhere – had a value for tenure.

<sup>7</sup> Units without a matching record in the administrative data were assigned a value for each of these variables to indicate that there was no record for the unit. A different value indicated there was a record for the unit, but the variable value was missing.



## National Estimates

The estimate of single attached units was reduced from 8.2% to 5.1%, compared to the ACS 2018 percentage of 5.9% (Table 7). The estimate of multi-units increased from 23.3% to 26.3%, compared to 26.6% from the ACS.

**Table 7.** Model-based structure type edit to AHS 2019 brought national estimates of single attached units and multi-units closer to American Community Survey 2018 estimates.

	<b>AHS 2019 National Estimates</b>		<b>ACS 2018</b>
	Millions (percentage)		
	<i>No Edit Applied</i>	<i>Edit Applied</i>	
Single Attached	11.4 (8.2)	7.2 (5.1)	8.1 (5.9)
Multi-units	32.5 (23.3)	36.8 (26.3)	36.8 (26.6)

## Metro Estimates

Among the 25 metro areas surveyed, the model-based structure type changes resulted in single attached estimates that were closer to the ACS 2018 estimate for 16 of them (Figure 3). For the metro multi-unit estimates, that number was 23 (Figure 4).

## Comparison with 2017 Edit

Figure 5 shows the 2019 AHS estimates for multi-units and single attached units (in dark green), for the top 15 metro areas and the nation, with the 2017 AHS estimates in gold. These estimates are shown as a percentage of their corresponding previous year ACS estimate. The points cluster more around  $y=100$  than  $x=100$ , indicating that both the 2017 edit and the 2019 edit brought most of these areas' estimates closer to the ACS estimate.

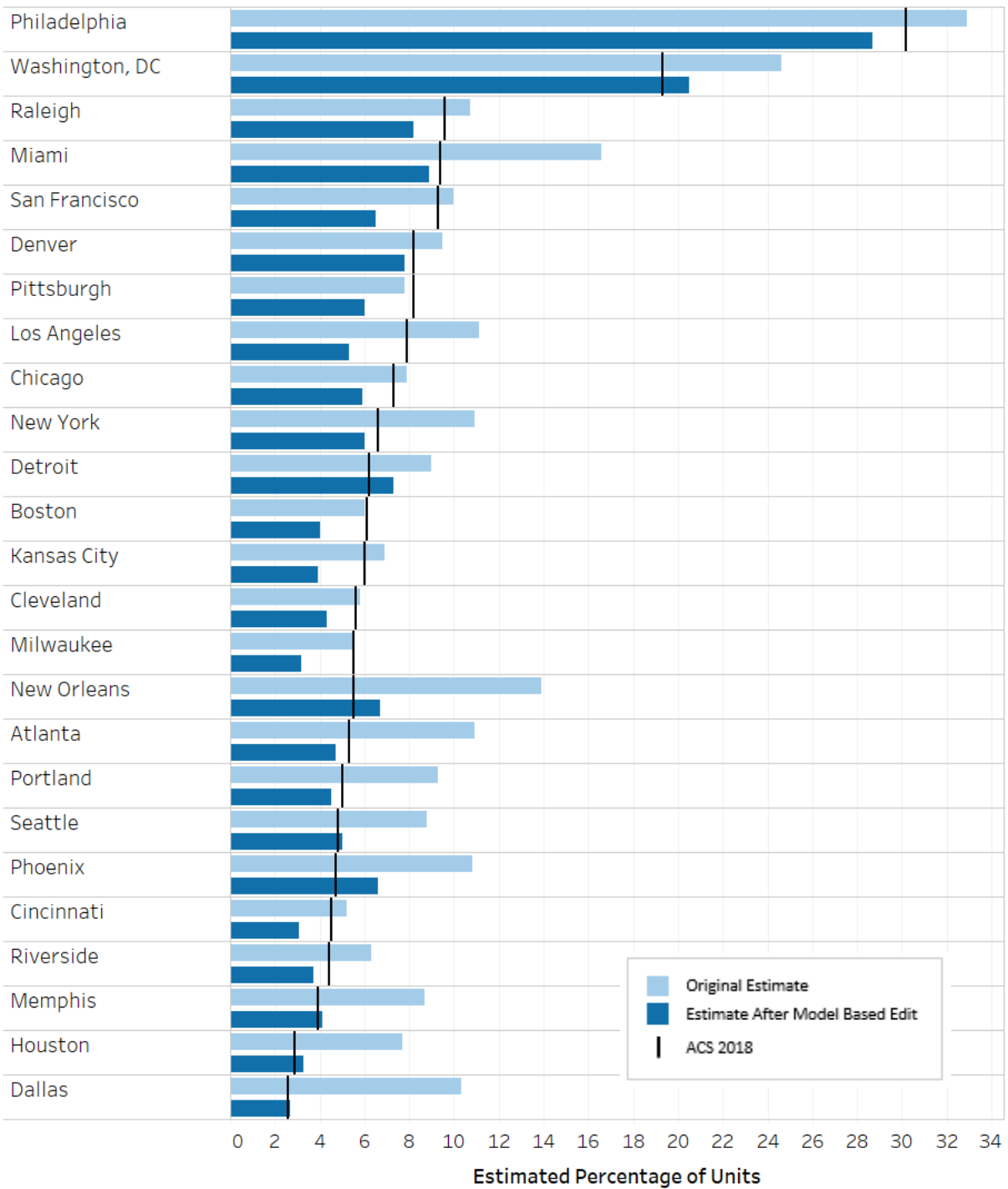
## Limitations of Approach

The somewhat high single attached misclassification rate of the models tested - above 30% for the best performing models - may result in a large proportion of true single attached units being recategorized to multi-units. This may result in an underestimate of the true number of single attached units, and an overestimate of multi-units.

Unit-level and item-level missingness were treated as answer categories in the model, rather than as entities that obscure the true distribution of a variable. This approach assumed that missingness has information about the response variable and is therefore not random, which appeared to be true for some variables selected for the model: item-level missingness in the administrative variables had been found to be associated with multi-units. The approach also assumed unit-level missings were not independent of the response, which Census did not verify.

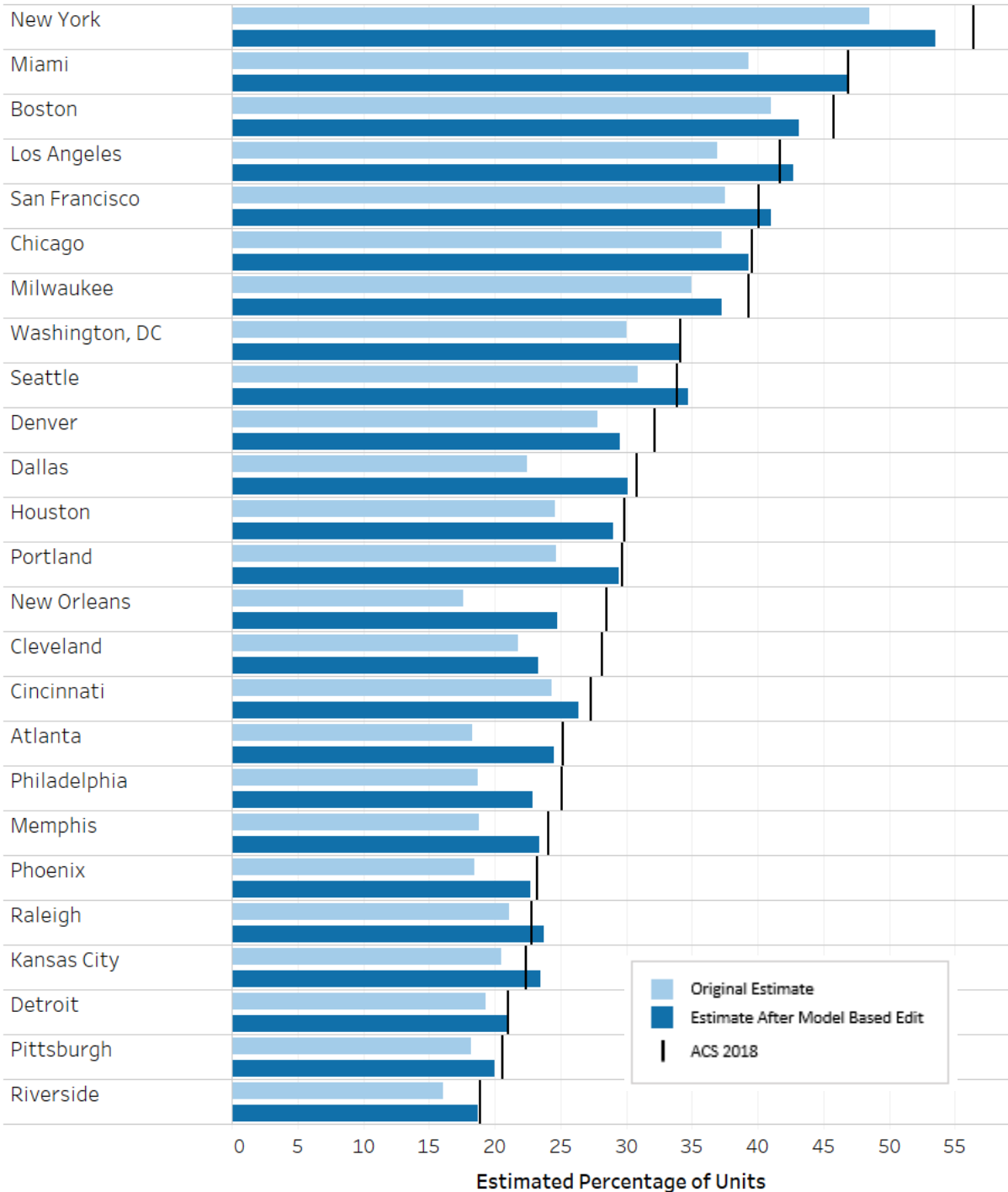
**Figure 3.** Single attached unit estimates in the 25 metropolitan areas surveyed, prior to and following model-based changes made to structure type. Reference line is the American Community Survey 2018 estimate for each metro area.

**Metropolitan Area Single Attached Unit Estimates Before and After Model Based Structure Type Edit, AHS 2019**



**Figure 4.** Multi-unit estimates in the 25 metropolitan areas surveyed, prior to and following model-based changes made to structure type. Reference line is the American Community Survey 2018 estimate for each metro area.

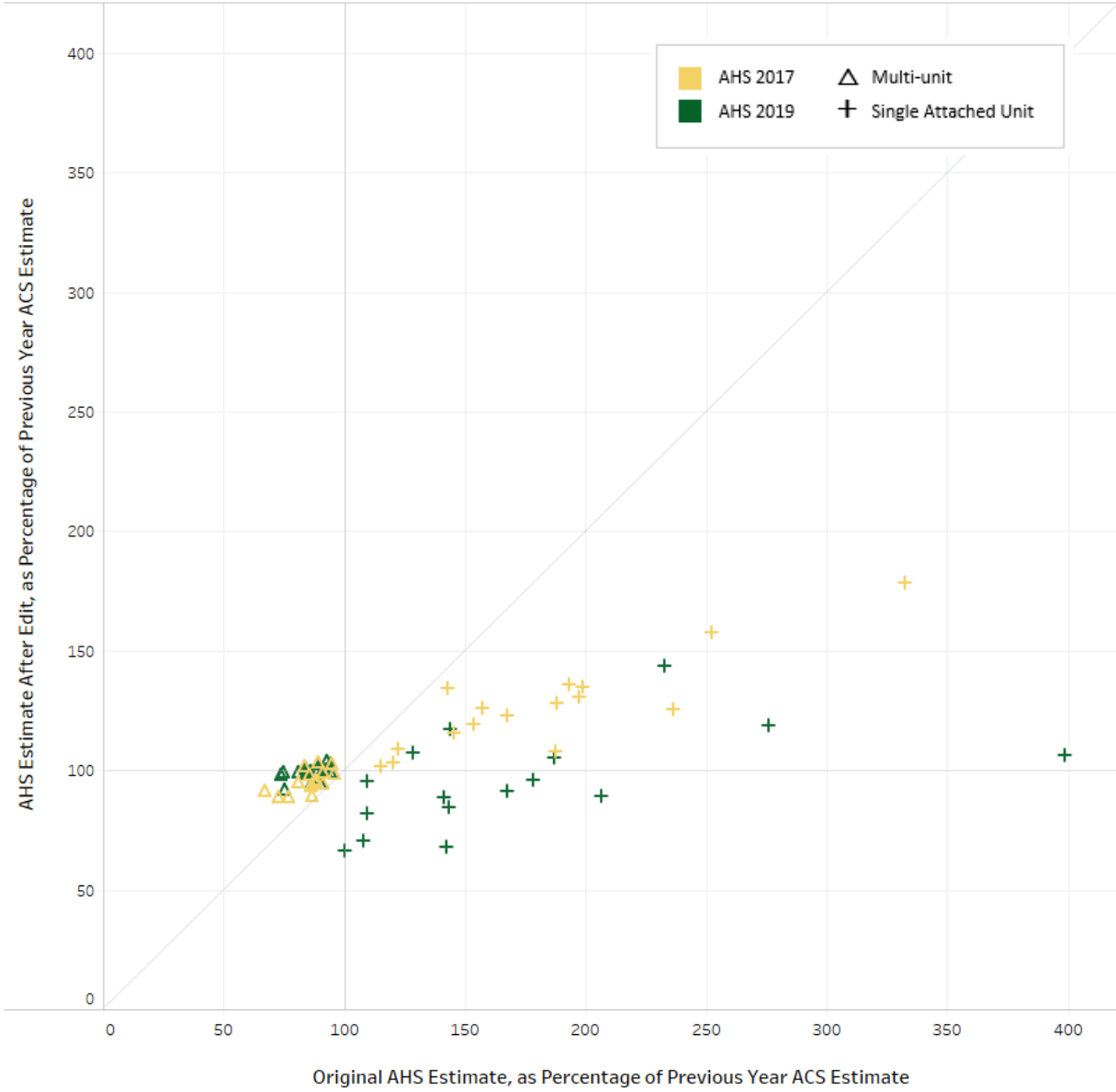
### Metropolitan Area Multi-Unit Estimates Before and After Model Based Structure Type Edit, AHS 2019



**Figure 5.** AHS multi-unit and single attached unit estimates for U.S. and top 15 largest metropolitan areas in 2017 and 2019, before and after structure type edit, as percentage of prior year ACS estimate. A point close to y=100 indicates the edit was effective in bringing the estimate more in line with the ACS estimate from the previous year for that geographical area. A point along the y=x line indicates the AHS estimate was changed very little by the

DRB Clearance Number - CBDRB-FY20-POP001-0175

edit. A point near x=100 shows the edit resulted in the AHS estimate being further from the previous year ACS estimate than it had been without the edit.



Figures 3 and 4 indicate that the percentage of units that are single attached varies widely across metro areas. Geography might be predictive of whether a unit is more likely to be single attached or a multi-unit, but there was no geography variable for the model selection mechanism to evaluate as a possible input variable.

### Effect of Structure Type Change on Other Variables

When a returning unit's structure type changed, the unit became eligible for questions it was not asked in 2019, but may have been asked in a previous wave. For the returning units whose structure type was changed by the model, Census obtained prior wave values for the questions whose instrument universe definitions used structure type, and that were asked only of either multi-units or single attached units. These values were merged in at an early data processing step along with the new structure type. The IUF variables affected were BLD, COMPLEX, ENTRYSYS, FOUNDTYPE, HOWBUY, KITEXCLU, LOTSIZE\_IUF, LOTSIZE, MGRONSITE, STORIES, and SUBDIV.

The questions these units were asked in 2019 but for which they were no longer in universe with the structure type change were blanked either as part of the prior wave data merge, or by processing rules.

## 6. 2021 AHS: Future Plans

In 2021, SHARAT and SHARFR will be asked of single attached units, and used to edit structure type in the instrument, as in 2019. ABUNIT1 and ABUNIT2 will be asked of single attached units and multi-units of all tenure types rather than only renters; they will not be re-asked of returning units that provided a response in 2019. If a single attached unit answers ABUNIT1 by reporting another unit above or below, the instrument will change the unit's structure type to multi-unit. ABUNIT2 responses may be considered as a possible input variable for a model to adjust structure type that could be implemented for the 2021 data, but they will not be used to edit the structure type in the instrument.