A Comparison of Training Modules for Administrative Records Use in Nonresponse Followup Operations: The 2010 Census and the American Community Survey*

Melissa C. Chow, Hubert P. Janicki, Mark J. Kutzbach, Lawrence F. Warren, and Moises Yi

Center for Economic Studies, U.S. Census Bureau,
4600 Silver Hill Road, Washington, D.C. 20233

Abstract

While modeling work in preparation for the 2020 Census has shown that administrative records can be predictive of Nonresponse Followup (NRFU) enumeration outcomes, there is scope to examine the robustness of the models by using more recent training data. The models deployed for workload removal from the 2015 and 2016 Census Tests were based on associations of the 2010 Census with administrative records. Training the same models with more recent data from the American Community Survey (ACS) can identify any changes in parameter associations over time that might reduce the accuracy of model predictions. Furthermore, more recent training data would allow for the incorporation of new administrative record sources not available in 2010. However, differences in ACS methodology and the smaller sample size may limit its applicability. This paper replicates earlier results and examines model predictions based on the ACS in comparison with NRFU outcomes. The evaluation consists of a comparison of predicted counts and household compositions with actual 2015 NRFU outcomes. The main findings are an overall validation of the methodology using independent data.

Key Words: Count imputation, administrative records, nonresponse, American Community Survey

1. Introduction

1.1 Background

In preparation for the next Decennial Census in 2020, the U.S. Census Bureau is seeking to reduce the costs associated with conducting the Census. During the 2010 Census, the largest contributor to cost was the Nonresponse Followup (NRFU) operation, which cost over $2 billion. The purpose of the NRFU operation was to obtain responses for those households and individuals who did not self-respond. This operation led to up to six visits by enumerators to each household. When planning for the 2020 Decennial Census, the U.S. Census Bureau searched for solutions to make the NRFU operation more efficient. One suggestion for reducing the number of NRFU personal visits is to use administrative record data to assess occupancy, manage workload, and for enumeration. More information on administative records use in NRFU operations is detailed in Mule and Keller (2014)

*Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is released.

and Keller (2016). The U.S. Census Bureau has administrative record data from various sources, including tax data from the Internal Revenue Service (IRS), other government agencies, and third-party records on individuals and households. Administrative records would be helpful in figuring out the composition and counts of the household, and could reduce the number of visits.

To analyze the data, two broad categories of models are used, one to identify occupied housing units for administrative record enumeration and one that identifies vacant units. The occupied category contains two models: (1) Household-Composition Model; (2) Person-Place Model. Meanwhile, the vacant category contains one model, the Vacancy Model. These models are based on the Household-Composition Model, the Person-Place Model, and the Vacancy Model described in Morris et al. (2016). However, in this analysis, instead of using the 2010 Census data to train the model, data from the 2014 American Community Survey (ACS) is used to predict data from the 2015 Census Test.

The ACS is a monthly survey conducted by the U.S. Census Bureau. The monthly samples are used to produce annually updated estimates for the same census tracts and block groups formerly surveyed via the Decennial Census long-form sample.[1] This analysis compares how the ACS data does at predicting the household compositions and counts for the 2015 Census Test, relative to the 2010 Census data. There are several reasons why implementing the models with ACS data may be worthwhile. First, the ACS provides an independent check of the training data. Second, collection of the ACS data is continuously ongoing and 2018 ACS data will be closer to the 2020 Census data than 2010 Census data in terms of date collected. Third, the ACS allows for evaluating administrative record sources that are not available for 2010. For example, the Census Bureau has Supplemental Nutrition Assistance Program (SNAP) data for certain states and years, but the data may not go as far back as 2010. Finally, the ACS only does household responses, so there is no proxy data.

There are a few drawbacks in using the ACS data. The main drawback is that the 2010 Census is all-encompassing and is meant to cover the entire population. The Decennial Census is more visible to the population, with extensive advertisement and promotion. On the other hand, the ACS only includes a sample of the population, with approximately 3.5 million households a year. Therefore, the set of ACS non-respondents may differ from the set of Census non-respondents. The Census data has a reference date of April 1 while the reference date for the ACS data, which is conducted in all months, is the date that the respondent answers the survey (regardless of which panel they are originally assigned to). Lastly, the NRFU operation differs between the 2010 Census and the ACS.

1.2 Outline

This paper proceeds as follows. We introduce the models and modifications specific to the ACS implementation in Section 2. We discuss the data sources and particularities of the ACS data in

[1] Additional information about the ACS can be found on: https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

Section 3. We discuss the results evaluating 2010 Census and ACS trained models in Section 4. We conclude in Section 5. We include detailed results for extensions in the Appendix.

2. Methodology

2.1 Model Structure

This analysis uses as a baseline the methodology described in Keller and Konicki (2016), which was implemented for the NRFU operation of the 2016 Census Test. The methodology targets occupied addresses for removal from NRFU after a single visit based on three modeling steps and the application of a decision rule. For that analysis, all three models were trained using 2010 Census responses at the address, or Master Address File ID (MAFID), level in combination with administrative records, as well as housing unit and neighborhood information.[2] The first "vacancy" (VAC) model explains the final status of a MAFID as occupied, vacant, or delete. The second "person-place" (PP) model explains the agreement of administrative records with responses using personal identifying information.[3] The third "household composition" (HHC) model explains the agreement of administrative records with the combination of adults and child responses, using AR and responses to evaluate age. These models are estimated independently on training data using a multinomial logit for the VAC and HHC models and a logit for the PP model. The parameter estimates are then used to score equivalent data for a NRFU operation, resulting in predicted probabilities for each model at each MAFID.

The determination of whether to remove a unit from the NRFU workload is based on the "distance function." This is in contrast to Keller and Konicki (2016) and Morris et al. (2016), who use linear programming techniques. The distance function was adopted in the 2016 Census Test, while linear programming was used in the 2015 Census Test. For more information on distance vs linear programming techniques, see U.S. Census Bureau (2017). For all MAFIDs that are determined to be occupied based on administrative records (from the VAC model), a distance, or d_{AR_occ}, is calculated as

$$d_{AR\_occ} = \sqrt{(1 - \hat{p}_{pp})^2 + (1 - \hat{p}_{hhc})^2}. \quad (1)$$

In 1, $\hat{p}_{pp}$ is the minimum predicted agreement from the PP model among all persons at a MAFID (the least certain person in that household). The value $\hat{p}_{hhc}$ is the predicted probability associated with the composition of observed administrative records for that MAFID [4]. Higher values of these predicted probabilities reduce the distance and indicate greater certainty in the administrative records. An independent optimization process determined a distance threshold, whereby all MAFIDs with a distance under the threshold are removed from the NRFU workload.

[2]For a detailed list of the covariates in each model, see Table 4 in Morris et al. (2016).

[3]The model explains agreement of administrative records and household responses on the assignment of a Protected Identification Key (PIK) for each person in a household and has been used to examine nonresponse to the 2010 Census (Brown, 2013)

[4]A similar distance function is also used to identify vacant households using the predicted probabilities from the VAC model.

2.2 ACS Implementation

The methodology as implemented using 2010 Census training data cannot be directly applied to 2014 ACS training data. Several challenges needed to be addressed. These challenges are due to differences in data collected regarding the mailings themselves and differences in the reference date and timing of followup operations of NRFU and ACS.

The first challenge concerns a difference in mailing data collected between the 2010 Decennial and the ACS. In particular, the 2010 Decennial collects information from the United States Postal Service on the reason for undeliverable-as-addressed (UAA) return code. This UAA reason is critical in establishing the vacancy status of an address and has a large impact on the types of housing addresses that are kept or removed from the fieldwork workload. UAA reasons include insufficient address (i.e. mail without a number or street), no such number, unclaimed, deceased, and vacant, among others. Unfortunately, the ACS collects whether the mailing has UAA status, but does not collect the reason for the UAA status as described by the U.S. Postal Service (USPS). ACS does not purchase UAA reason data from the USPS because they are not needed due to differences in data collection methodology. Nevertheless, the consequence of missing UAA reason is that we are unable to estimate the VAC model with the same precision as the 2010 Census. The final status flags used in this analysis (occupied, vacant, or delete) are taken from a model estimated on 2010 Census rather than 2014 ACS data. This *hybrid* approach still allows us to use the more recent information on household composition and counts from the ACS, but takes the best available information from the 2010 Census to inform the vacancy model. Since all models are estimated before any addresses are removed for vacancy or occupied removal, the training source for the Vacancy model does not affect the estimates for the HHC and PP models. The list of addresses removed by the Vacancy model are identical for both Census-trained and ACS-trained occupied removal, giving the same set of potential addresses to be removed by the occupied-removal process. This setup allows for better comparison of the occupied-removal outcomes using 2010 Census and ACS. The difference in availability of detailed UAA status between 2010 Census and ACS is also reflected in the covariates used in the PP and HHC model. The detailed UAA reason codes are covariates in the Census-trained models, while only the presence of UAA is used as a covariate in the ACS verions of the HHC and PP models.

The second challenge concerns a difference in reference dates and followup operations. The reference period of the 2010 Decennial is April 1 regardless of when the respondent fills out the form or when a non-response interview takes place. The NRFU operations begin in May with in-person visits. In contrast, the reference period of the ACS response is the date that the response was obtained, regardless of the month chosen for the sampling frame. ACS fieldwork operations are also quite different from the 2010 Census.[5] In particular, fieldwork begins with a telephone stage one month after the initial mailing and an in-person component that begins two months after the initial mailing. This timing difference presents a unique problem since responses obtained can reference a time period months after the initial mailing. One consequence of this problem is that

[5]The two followup operations are known as Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Personal Interviewing (CAPI)

potential respondents can move in and out of an address unit in the time between an initial mailing and a fieldwork interview leading to differences in vacancy, count, and composition of an address.

We attempt to gauge whether reference date inconsistency impacts the accuracy of model predictions by varying the timing of ACS training data used to estimate the HHC and PP model. Specifically, we estimate the HHC and PP models using *all* ACS addresses as our baseline. We then restrict our sample to respondents with an April reference date only. That is, we restrict our sample to mail-in responses that were filled out in April or when fieldwork was completed in April (as defined in the RDATE variable). We conduct a similar evaluation with other months. Finally, we exclude addresses that were sub-sampled out of fieldwork operations before the in-person follow-up visit (PI_ST variable status code 100).

2.3 Evaluation Framework

Our evaluation strategy is to compare the outcomes of using the ACS-trained and Census-trained models to select different removal samples of NRFU addresses. For each sample of removed addresses, we compare the AR-determined counts and household compositions with the counts and compositions reported in NRFU for the 2015 Census Test. This strategy follows the same methodology used by Mulry et al. (2016) in comparing the 2010 Census-based models to NRFU operations in the 2015 Census Test. Given the limitations of ACS for determining the vacant status of an address, we focus on explaining the count and composition of occupied units. Note that since covariates and sample size differ between the ACS and Census-based models, there is not a good direct comparison of model coefficients or goodness-of-fit. Our means of comparison between ACS and Census-based predictions is in the match rates of the addresses removed.

We investigate to what extent administrative records agree in household count and composition for NRFU cases that were removed from followup using both the 2010 Census and 2014 ACS data. To what extent does the model accurately predict household counts and compositions? In the cases that it does not accurately capture household characteristics, what is the magnitude of the disagreement in counts?

We then compare agreement of ACS-based predictions with those based on 2010 Census data. In particular, we want to evaluate to what extent the two models agree on the addresses to be removed from workload. If there are addresses that are not common to both models, we will examine address characteristics and find out if administrative records and 2015 NRFU responses agree in count and composition. Since the distribution of predicted probabilities and corresponding distances may not be the same using two different training samples, we evaluate the performance of ACS-based predictions by removing a set number of observations rather than using a threshold distance cutoff.

To further study the effects of variation in training data on predictions, we compare the sensitivity of the removal samples and match rates when the scope of the ACS training data is varied. Specifically, we address how well ACS-based models perform when data are restricted to Arizona only (state of 2015 Census Test) or April respondents only, as well as the sampling panel for each month from February through July, 2014.

Our evaluation framework can be summarized as follows:

1. We estimate the VAC model using 2010 Census data and the PP and HHC models using both 2010 Census and 2014 ACS data.
2. We then eliminate likely vacant addresses via the vacancy model as described in Mulry et al. (2016).
3. We use the PP and HHC models to calculate the distance function values for the remaining addresses using both Census and ACS-trained predictions, restricting the households to evaluate for removal to those that have an occupancy count of 6 individuals or less and fit within one of 6 household composition types[6] in administrative records.
4. We select the 3,400 households with the smallest calculated distances using the 2010 Census-trained models and the 3,400 households with the smallest distances based on 2014 ACS-trained models.
5. We evaluate the performance of the workload removal using this procedure by comparing the percentage of addresses for which administrative record counts or compositions matched the actual household population counts and compositions reported during 2015 Census Test NRFU.

3. Data

We construct two training datasets and one evaluation dataset. The 2010 Census and 2014 ACS are each used to develop a training set, while 2015 Census Test data are used for evaluation. Contemporaneous administrative records as well as neighborhood and address information are used in all three datasets.

3.1 Administrative Records

We use several sources of administrative records. Internal Revenue Service (IRS) sources are composed of Individual Tax Returns (1040) filed in tax years 2009, 2013, 2014 and weeks 4-17 in tax years 2010, 2014, and 2015. We also use IRS Information Returns 1099 for 2010, 2014, and 2015. In addition, we use Medicare enrollment data from the Center for Medicaid and Medicare Services (CMS) and Indian Health Services Patient Database.

Finally, we use information from the TARGUS database. This is a commercial data source that provides person verification. We also make use of data from the United States Postal Service (USPS) to inform the model with undeliverable-as-addressed (UAA) flag and reason (for 2010 model only). Data from these administrative records sources are matched with person and place observations in the 2010 Census, 2014 ACS, and 2015 Census Test when possible.

[6]These household compositions are 1, 2, or 3 adults, with or without children. See Section 4 and Mulry et al. (2016) for details.

3.2 2010 Census

We use the 2010 Census as the baseline dataset to which we make our training comparisons. We restrict our use of 2010 Census data to the universe of NRFU cases in Arizona only. The restriction to Arizona cases coincides with state chosen for the 2015 Census Test. We use respondent age variables to construct counts and household composition (by age) variables at each address. We augment the 2010 Census data with additional variables from the Master Address File (MAF) to obtain address characteristics of residences (such as the type of housing unit). Finally, the dataset is then linked with administrative records using MAFID and PIK variables.

3.3 American Community Survey

We use the 2014 American Community Survey (ACS) as our primary source of data for model estimation outside of 2010. The ACS is a nationwide survey designed to provide communities with a fresh look on how they are changing. The ACS replaced the decennial long form in 2010 by collecting long form type information throughout the decade. Data used in this report is based on the initial sample of 2014 ACS that includes respondents and non-respondents. In particular, the dataset includes non-respondents that were sub-sampled out due to unmailable or non-responding addresses that were not referred to a telephone-based or in-person followup. See U.S. Census Bureau (2014) for a detailed description of sampling methodology. The ACS data used for this analysis was not "swapped," a disclosure limitation designed to protect confidentiality of certain at-risk households that is present in public-use data (Lauger et al. (2014)).[7]

For our analysis, we make use of five internal ACS files. The *control* file contains data on sampling frame, mailing and CATI/CAPI outcome codes that are needed in our analysis to distinguish between households that were sent to followup and those that were not. The *household* file contains household level variables needed to estimate the household composition model. The *person* file contains person-level variables needed to estimate the PP model. The *address* file contains the necessary crosswalk to obtain address-level identifiers (MAFID) needed to merge in administrative records. Finally, we also require an extract from the Master Address File (MAF) that serves as the original sampling frame of addresses. This dataset is needed to obtain address characteristics at the time sampling is conducted. These data sets are merged by internal ACS identifiers when available. The merged ACS dataset is then linked with administrative records using MAFID and PIK variables.

3.4 2015 Census Test

The 2015 Census Test took place between April 1, 2015 and August 14, 2015. The purpose of the test was to evaluate methods used to reduce fieldwork and data collection. The test site chosen included several areas within Maricopa County in Arizona. See Mulry et al. (2016) for more

[7]No confidential information is released. Pre-swapped ACS data are used as an input into the estimation of the three regression models. No ACS data are tabulated and no model estimates based solely on ACS data are reported.

details on the location and methodology. We restrict our NRFU universe to the *control panel* that mimicked the followup methodology of the 2010 Census. We used respondent age variables to construct counts and household composition (by age) variables at each address. We augmented the Census Test data with additional variables from the Master Address File (MAF) to obtain address characteristics at time of sampling. Finally, the dataset is then linked with administrative records using MAFID and PIK variables.

4. Results

In this section, we first establish a baseline version of the AR model trained using the 2010 Census and evaluated for the 2015 Census Test. Next, we implement the same model estimation and evaluation for 2014 ACS with a sample including all records except those subsampled out of the ACS NRFU. Last, we implement several extensions, considering alternate ACS training samples. The evaluation framework, in terms of comparing counts and compositions with the 2015 Census Test, is styled after Mulry et al. (2016).

4.1 Census 2010 Training

We first conduct our evaluation procedure as described in Section 2.3 based on models using Decennial Census data. The 3,400 units with the smallest value of the distance function (with a high likelihood of concordance with administrative records) are enumerated using administrative records and evaluated against NRFU results of the 2015 Census Test. 3,400 units were chosen as the size of the removal sample as this corresponds with removal of 10 percent of the NRFU workload. Within this sample of addresses, the goal is to evaluate the success of the modeling process in identifying records that can be accurately enumerated via administrative records.

We compare the results of AR enumeration of these addresses with the actual responses collected during NRFU for the 2015 Census Test. In Table 1, we start by presenting the comparison of population counts between AR and NRFU. We show figures both for the full evaluation sample, and for each type of household composition as determined by administrative records. Column 2 shows the number of households in each household category, while columns 3-5 show what percentage of households in each category had a higher, equal, or lower population count in the administrative records relative to the fieldwork records.[8]

We find that the household counts from the AR enumeration coincide with NRFU counts for 56.6 percent of addresses, but that there is generally an AR overcount for larger AR households. Of household compositions determined via administrative records, households with one adult and at least one child were most likely to be different from the NRFU household count, with only a 32.5 percent match rate. Administrative records and NRFU household counts match for 62.3 percent of cases with 2 adults and no children.

To illustrate the size of the population count differences introduced by using AR enumeration, Table 2 shows the distribution of count discrepancies by the magnitude of the differences in AR

[8]The last column shows the percentage of households with an unknown fieldwork population count.

Table 1: Population Count Comparison by AR Household Composition- Full Census 2010 Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	1107	7.2	57.5	32.2	3.2
1 adult, 1+ child	166	29.5	32.5	36.1	1.8
2 adults, 0 child	964	18.9	62.3	15.6	3.2
2 adults, 1+ child	982	24.6	55.9	17.7	1.7
3 adults, 0 child	36	41.7	44.4	13.9	0.0
3 adults, 1+ child	145	42.1	47.6	8.3	2.1
Total	3400	18.5	56.6	22.3	2.6

relative to NRFU responses. 6.7 percent of administrative records overcount household population by 2 or more individuals. 11.8 percent of administrative records overcount household population by one individual. The distribution of undercounts follows a similar distribution, as can be seen in columns 6 and 7. Importantly, the symmetry in the distribution of over and undercounts suggests that, on an aggregate level, AR enumeration would avoid over and undercounting of the population. Overall, administrative record enumerations match household population counts in the Census Test within one individual for 83.4 percent of addresses.

Table 2: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full Census 2010 Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	3400	6.7	11.8	56.6	15.0	7.3	2.6

Note: Table should be read as AR count relative to NRFU count.

Table 3 presents another comparison between the AR and NRFU records, this time focusing on the household composition classification instead of household population counts. Each row in Table 3 corresponds to a different AR household composition (as determined by administrative records), and each column corresponds to the household composition assigned during the NRFU followup. Each cell presents the share of observations for a given AR type that was classified under each different NRFU composition.[9]

The dominant diagonal values for each composition indicate that AR coincides with reported

[9]Each cell contains a row percentage, so the figures in each row add up to 100.

Table 3: Household Composition Comparison AR vs. NRFU - Full Census 2010 Sample

AR	NRFU								
	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	49.7	2.7	15.9	4.8	2.3	0.8	15.3	1.3	7.2
1 adult, 1+ child	11.4	41.6	4.8	16.3	0.6	0.6	16.9	2.4	5.4
2 adults, 0 child	11.2	0.6	55.6	4.6	5.7	1.7	13.3	2.2	5.2
2 adults, 1+ child	4.1	6.8	3.4	60.9	0.9	4.7	13.2	1.7	4.3
3 adults, 0 child	5.6	0.0	30.6	0.0	41.7	0.0	11.1	11.1	0.0
3 adults, 1+ child	2.8	6.9	2.8	20.7	4.1	47.6	9.7	3.4	2.1
Total	21.3	5.4	22.6	22.1	3.3	4.1	13.9	1.9	5.4

household compositions for a majority of cases. Of cases where AR compositions do not coincide with the NRFU followup, some patterns emerge. A large proportion of households classified as “3 adults, no children" are actually 2-adult households with no children. Similarly, many households classified as single-adult households actually contained 2 adults, and vice versa. Missing or unknown age in the non-response followup is also an issue for comparing household composition, as a missing age for one individual in the household results in the inability to classify the household’s composition. Smaller household compositions in AR were also more likely to be vacant than larger household compositions. It is worth noting that the AR compositions with the worst match rate are those of 1-adult and 1+ child, and 3-adults and no children. These compositions are also the least common households in our sample of 3,400 households. The overall match rate of household compositions is 54 percent.

4.2 ACS Training Baseline

Table 4 follows the same structure as Table 1 except it uses the ACS sample for training. The observations included in this table are the 3,400 that are removed from NRFU after the distance function calculation. Overall, the ACS baseline results are similar to those from the 2010 Census. In this subsection, we describe the ACS results and some differences relative to the 2010 Census results. The differences we describe here are mostly of low magnitude. We have not computed uncertainty measures and do not make statements regarding the statistical significance of differences.[10] Again we stress that due to model differences due to data availability in the training datasets, direct comparison of model coefficients would be misleading. We focus instead on a comparison of match rates between ACS and Census-based predictions of addresses removed from the NRFU workload.

Compared to the results from the 2010 Census sample, the ACS sample includes a larger number of households with one or more adults and zero children. When comparing the two samples, the ACS has a greater percentage of having an equal household composition for units with two or more adults and one or more children. Across all compositions, the ACS has a higher percentage of being equal to the household counts reported via NRFU than the 2010 Census sample.

Table 4: Population Count Comparison by AR Household Composition- Full ACS Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	1240	10.1	57.8	29.0	3.1
1 adult, 1+ child	50	36.0	28.0	34.0	2.0
2 adults, 0 child	1112	21.3	61.9	13.2	3.6
2 adults, 1+ child	917	22.4	56.8	19.0	1.9
3 adults, 0 child	38	47.4	36.8	13.2	2.6
3 adults, 1+ child	43	30.2	60.5	7.0	2.3
Total	3400	18.1	58.2	20.7	2.9

Table 5 is analogous to Table 2, but for the full ACS sample. Here, the ACS results have a higher percentage match based on population counts. However, they also have a slightly higher percentage of having one fewer person in the household and a slightly higher percentage of NRFU count unknown compared to the 2010 Census sample.

Likewise, Table 6, based on ACS training, is analogous to Table 3, based on 2010 Census training. Again, there is a "dominant diagonal", with AR household compositions being more likely to correspond to an identical response household than any other type. The average correspondence

[10]We leave this to future work. However, we are reasonably confident that the addition of point estimate uncertainty into the analysis will have a small effect. In particular, note that selecting training samples by month, NRFU status, and state, has little effect on the overall match rates in count and compositions. This suggests that the results are invariant to some inclusion of point estimate variability.

Table 5: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full ACS Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	3400	5.6	12.5	58.2	14.1	6.6	2.9

Note: Table should be read as AR count relative to NRFU count.

rate is 54.4 percent with ACS training, compared to 54 percent in 2010 Census training. However, the magnitude varies across types, with some household types having greater correspondence in Table 3 and others in Table 6. In general, the 2010 Census training model agrees more often for the AR households with no children, while the ACS trained model agrees more often for households with children.

Table 6: Household Composition Comparison AR vs. NRFU - Full ACS Sample

AR	NRFU 1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	48.6	1.9	15.9	3.6	2.2	0.6	15.8	1.2	10.1
1 adult, 1+ child	10.0	42.0	4.0	6.0	6.0	4.0	12.0	2.0	14.0
2 adults, 0 child	13.5	1.3	54.3	4.3	4.4	1.3	12.7	2.2	6.0
2 adults, 1+ child	3.4	6.2	3.2	63.4	0.9	4.3	13.5	1.6	3.6
3 adults, 0 child	10.5	0.0	34.2	0.0	31.6	0.0	10.5	13.2	0.0
3 adults, 1+ child	2.3	4.7	0.0	11.6	0.0	62.8	9.3	4.7	4.7
Total	23.4	3.5	24.9	20.1	2.9	2.6	14.0	1.9	6.9

4.3 Comparison of Training Modules

Table 7: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	54.0	56.6
Baseline - ACS	54.4	58.2

Having discussed the evaluation results for both the 2010 Census and ACS training modules,

we now present comparative analyses to highlight the similarities and differences of the results. A comparison of model predictions for the ACS trained and 2010 Census trained models finds overall similarity in the accuracy of predictions for count and household composition. See Table 7. The baseline and most complete ACS training sample matches the NRFU responses in count with a rate of 58.2 percent compared to 56.6 percent for 2010 Census training. Likewise, for household composition, the respective rates were 54.4 and 54.0 percent. These results suggest that the mixed approach, using 2010 Census to evaluate vacancy and ACS for the PP and HC models, minimized the impact of not having detailed UAA codes. Furthermore, the year-round sampling and smaller sample size do not seem to have resulted in worse, overall accuracy for the ACS trained model. While the ACS trained model achieves a slightly higher agreement rate, we do not regard these differences to be of sufficient magnitude to conclude that the ACS is actually a superior training module. Rather, these results suggest that the ACS would be an appropriate substitute for evaluating and updating the model and incorporating new administrative records.

Tables 8 and 9 provide insight into the degree of overlap in the workload removal from the 2010 Census and ACS trained models, respectively. Each table ranks the 3,400 records removed from workload by ascending deciles of the distance function value, with the first decile being the records removed with the greatest degree of confidence. We provide the degree of overlap with the training data from the alternate training module, by decile. For both modules, overlap is near 100 percent in the first decile and close to 50 percent in the tenth decile. The average overlap of about three quarters explains the similar match rates of the two training modules and suggests a high degree of agreement in which records to remove.

Table 8: Sample Overlap - Census vs. ACS (by Census distance)

Decile of Census Distance	% present in ACS sample	Cut-off Census distances
1 - Highest precision	100.0	0.494
2	99.1	0.552
3	96.8	0.596
4	96.5	0.638
5	84.1	0.678
6	76.8	0.714
7	68.0	0.746
8	64.9	0.773
9	52.1	0.796
10 - Lowest precision	47.9	0.820

Figures 1 and 2 demonstrate the tradeoff of the quantity of records removed and the marginal disagreement rate in household composition and population counts, respectively. Each figure illustrates this tradeoff for both the 2010 Census and ACS trained models. The horizontal axis lists bins of the distance rank for each module, with 20 bins encompassing the 6,800 records with the lowest distance scores (bins 1 through 10 contain the 3,400 units in the removal sample). As with

Table 9: Sample Overlap - Census vs. ACS (by ACS distance)

Decile of ACS Distance	% present in Census sample	Cut-off ACS distances
1 - Highest precision	99.1	0.316
2	98.2	0.370
3	94.7	0.414
4	87.1	0.454
5	74.4	0.492
6	84.1	0.547
7	71.2	0.590
8	63.2	0.630
9	64.4	0.667
10 - Lowest precision	49.7	0.702

the decile bins in Tables 8 and 9, the 20th bin includes the records removed with the least confidence. For each bin, the vertical axis gives the disagreement rate, constructed as one minus the agreement, or match rate, from Table 7. These figures plot the tradeoff of less agreement associated with removing a greater quantity of records. The Census 2010 and ACS trained modules appear to have a similar tradeoff for both count and composition across the full range of the distance function presented here.

Figure 1: Household Composition Disagreement Rates and Distances (Census vs. ACS)

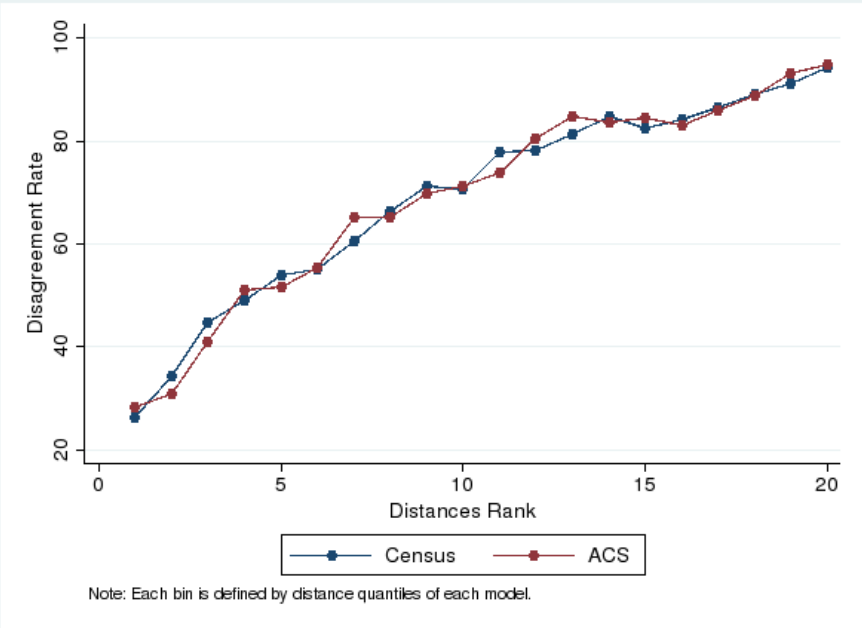
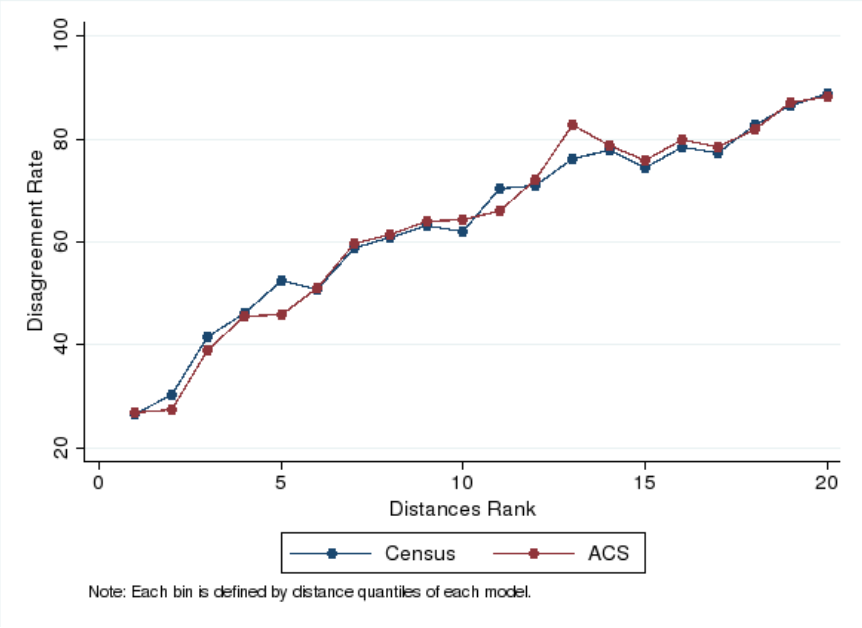


Figure 2: Household Population Count Disagreement Rates and Distances (Census vs. ACS)



4.4 ACS Training Extensions

In this section, we summarize the overall match rates in counts and household composition by various definitions of ACS data used to estimate our predictive models and compare them to our baseline 2010 Census dataset. In Table 10, we compare the ability of each model to match the counts and household composition as predicted by household responses. By varying the input dataset used to create predicted probabilities of count and composition, we can evaluate the ability of each model to correctly predict the household response.

Table 10: Comparison of Matches in Household Composition

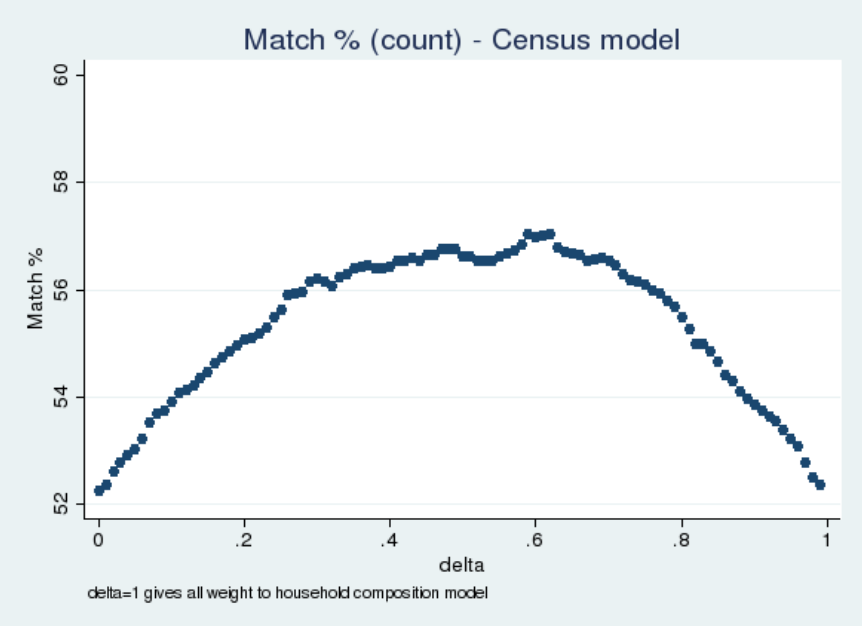
Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	54.0	56.6
Baseline - ACS	54.4	58.2
ACS - Arizona	54.9	59.1
ACS - April panel (responses)	53.9	58.4
ACS - February panel	53.8	58.1
ACS - March panel	54.0	58.4
ACS - April panel	54.1	58.5
ACS - May panel	54.3	58.6
ACS - June panel	54.4	58.7
ACS - July panel	54.4	58.6
ACS - NRFU	54.4	58.2

Note: NRFU training based on ACS MAFIDS where telephone and in-person follow-ups (CATI and CAPI) were conducted due to non-response of the mail-in form.

The first two rows of Table 10 present the composition category and count match rates found using a model estimated using 2010 Census data and ACS data. These two rows are the topic of the previous section and are listed here for reference. They show that a model estimated on ACS data (baseline) is comparable to one estimated on 2010 Census data with household composition match rates of 54.0 percent and 54.4 percent and count match rates of 56.6 percent and 58.2 percent.

Notice that the baseline ACS model uses all available responses regardless of residence or time of response. As discussed in Section 2.2, there are differences in time and geographical scope of the ACS data that are available for model estimation. The key feature seen from Table 10 is that changes in ACS data used to estimate the models has a negligible effect on the accuracy of counts and household composition. For example, restricting the ACS data to respondents from Arizona (row 3) only *increases* the category match rate from 54.4 percent in the baseline case to 54.7 percent. In comparison, restricting ACS data used in model estimation to respondents that provided a response in April (either self-respondent or through NRFU) decreases the category match rate from 54.5 percent to 53.9 percent. Rows 5 through 10 further explore the accuracy of models trained on different months of ACS responses. While accuracy is fairly constant, we note that it tends to improve for later months, even those after April. The worst accuracy, 53.8 and 58.1

Figure 3: Household Unit Count Match Rates and Distance Weights (Census Model)



percent match rate in category and count respectively is for the February panel, compared to 54.4 and 58.6 percent in July.

4.5 Sensitivity Analysis of Distance Function Weighting

In this section, we examine the sensitivity of matching results to the weighting of the PP and HC models in the distance function. Equation 1 assumes an equal weighting of the predicted probabilities for each model. We consider the full range of alternate weights, writing the function as

$$d_{AR\_occ} = \sqrt{2\delta(1-\hat{p}_{hhc})^2 + 2(1-\delta)(1-\hat{p}_{pp})^2} \tag{2}$$

for $\delta \in [0, 1]$, which give greater weight to the household-composition model as δ approaches 1. For each of 100 values of δ, we evaluate the distance function, or Equation 2, and select the 3,400 units to remove with the lowest weighted distances. For both the 2010 Census and ACS trained models, we calculate the person-place and household-composition match rates for each scenario. We graph the match rates by δ in Figures 3, 4, 5, and 6.

Each of the figures demonstrates the value of the hybrid model, which is apparent from the inverted "U" shape. One difference of the 2010 Census and ACS trained models is that the former is optimal closer to an even weight split, while the latter favors the PP model.

Figure 4: Household Composition Match Rates and Distance Weights (Census Model)

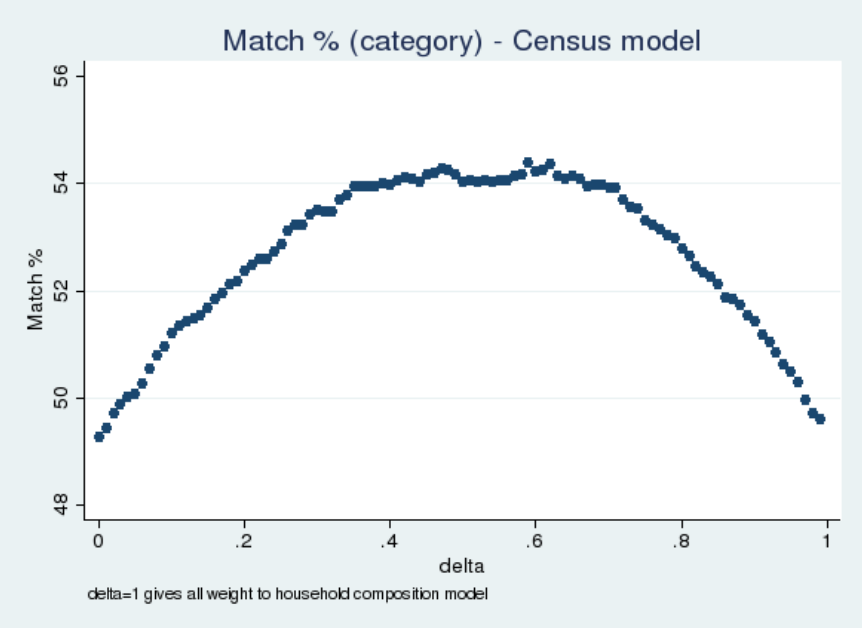


Figure 5: Household Unit Count Match Rates and Distance Weights (ACS Model)

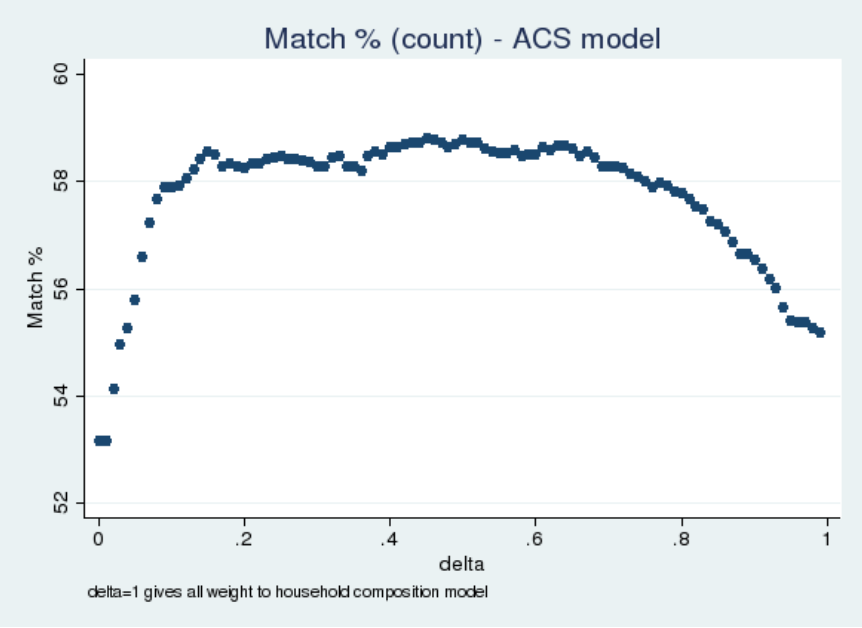
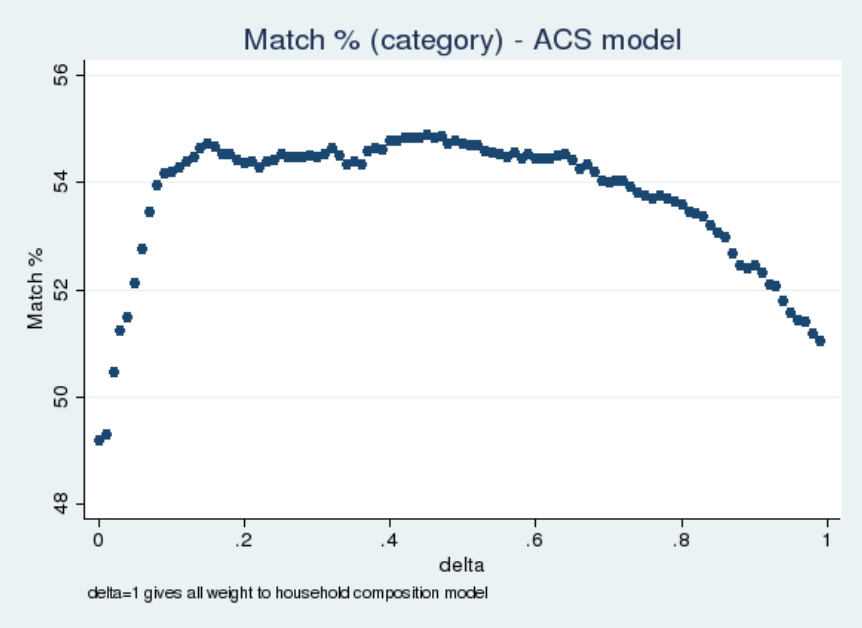


Figure 6: Household Composition Match Rates and Distance Weights (ACS Model)



4.6 2016 Census Test Extensions

For a robustness check, the analysis using ACS data was repeated using data from the 2016 Census Test. The 2016 test focused on the cities of Los Angeles, California and Houston, Texas, and the detailed results are presented in Appendices B and C, respectively. The 2016 test differed from 2015 by having a larger population and by using a different methodology for retaining an evaluation sample. In terms of household composition, compared with 2015, the match rate in 2016 was lower for both the 2010 Census and ACS trained models. In terms of count, compared with 2015, the match rate in 2016 was higher in Los Angeles, but lower in Houston, for both the 2010 Census and ACS trained models. The 2010 Census and ACS match rates closely tracked each other across these evaluations. Overall, these extensions support our main conclusion that the ACS training module performs similarly to a 2010 Census training module.

4.7 2014 ACS Extension and SNAP Analysis

To test the utility of ACS data as a more contemporary source of training data, we use a modified version of the 2015 and 2016 Census Test analysis to evaluate the addition of supplemental data in Appendix D. ACS data may be useful as an option for analysis in cases where data is unavailable back to 2010, or coverage does not exist in a Census Test location and year. We consider incorporating enrollment data from one such data source, the Supplemental Nutrition Assistance Program (SNAP), to our list of covariates. SNAP enrollment data are not widely available back to 2010,

which would be necessary to conduct analysis using Decennial Census data. Additionally, SNAP information is not available in any of the state-year combinations of the 2015 and 2016 Census Tests. We compare the results of training using ACS data from 2014 with and without SNAP information for the state of Illinois. The primary change in the analysis is the need to use a random subsample of addresses from ACS for evaluation as well as training (because there is no Census Test). Although the improvement in results from using SNAP enrollment data is minimal, this exercise suggests that using ACS for both training and evaluation is a plausible method to test the utility of potential additional data when coverage or availability is limited.

5. Conclusion

This evaluation shows that the ACS performs comparably with the 2010 Census as a source of training data for AR models used for 2015 NRFU, with similar count and composition predictions and a high degree of overlap in the record sets selected for removal. The concerns of year-round sampling and a smaller sample size seem to have had minimal effect on the model accuracy, though models estimated based on later months and those using larger ACS samples tended to be more accurate. In summary, these results indicate that associations of AR and followup responses have not changed appreciably from 2010 to 2014 and that AR model predictions are not especially sensitive to the differences of Census and ACS fieldwork. Expanding our analysis to 2016 Census Test data and model estimation using 2015 ACS does not alter any of the conclusions drawn with earlier data.

The ACS training module presents several opportunities for further development. For example, addressing the lack of UAA data in the ACS training data may improve model accuracy. In addition, methods for boosting the specificity of ACS model estimates while maintaining sample size, such as developing synthetic training sets targeted for specific geographies, may further improve accuracy.

Acknowledgements

This research was conducted in coordination with the U.S. Census Bureau's Decennial Statistical Studies Division (DSSD) and builds on previous work. We thank Andrew Keller, Scott Konicki, and Thomas Mule for assistance with understanding the 2020 administrative records modeling methodologies and goals, sharing the latest research code, preparing data extracts, and for comments on this report. We thank the entire administrative records team, also including Michael Ikeda, Ingrid Kjeldgaard, and Darcy Morris for feedback and comments. We thank David Raglin and Larry Bates for assistance with obtaining the necessary American Community Survey and Master Address File datasets. We thank Erika McEntarfer and Shawn Klimek for ongoing support with coordinating the Center for Economic Studies (CES) team engaged in this project.

References

Brown, J. D. Synthesizing Numerous Pre-Existing Data Sources to Accurately Enumerate Nonresponding Decennial Housing Units. Technical report, U.S Census Bureau, 2013.

Keller, A. Imputation Research for the 2020 Census. *Statistical Journal of the International Association of Official Statistics*, 32:189–198, 2016.

Keller, A. and Konicki, S. Using 2010 Census Coverage Measurement Results to Better Understand Possible Administrative Records Incorporation in the Decennial Census. *JSM Proceedings*, Survey Research Methods Section, 2016.

Lauger, A., Wisniewski, B., and McKenna, L. Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research. Center for Disclosure Avoidance Research Report 2014-02, U.S. Census Bureau, 2014.

Morris, D. S., Keller, A., and Clark, B. An Approach for Using Administrative Records to Reduce Contacts in the 2020 Census. *Statistical Journal of the International Association of Official Statistics*, 32:177–188, 2016.

Mule, T. and Keller, A. Using Administrative Records to Reduce Nonresponse Followup Operations. *JSM Proceedings*, Survey Research Methods Section, 2014.

Mulry, M., Clark, B., and Mule, T. Using the 2015 Census Test Evaluation Followup to Compare the Nonresponse Followup with Administrative Records. *JSM Proceedings*, Survey Research Methods Section, 2016.

U.S. Census Bureau. American Community Survey Design and Methodology. Technical report, U.S. Census Bureau, 2014.

U.S. Census Bureau. Administrative Records Modeling Update for the Census Scientific Advisory Committee. Technical report, U.S. Census Bureau, 2017.

A. Appendix: ACS Training Extensions

In this appendix, we discuss the impact of alternative ACS data definitions used to estimate our predictive model. The definitions discussed here focus on limiting ACS data to Arizona respondents, national respondents sampled in April, and national respondents that responded in April.

A.1 ACS - Arizona Only

In Table A.1 we display model results when coefficients are estimated using ACS respondents from Arizona. The purpose of this experiment is to see whether restricting our ACS data from all respondents to those from Arizona increases the ability of the model to more accurately predict household composition in the 2015 Census Test that took place in Maricopa County, Arizona. The most frequent category of the 3400 chosen addresses is 1 adult and 0 children. The model predicts the person count for this composition category accurately 57.5 percent of cases. Table A.2 presents a comparison of aggregate counts, which agree in 59.1 percent of cases.

Table A.1: Population Count Comparison by AR Household Composition- Arizona Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	1378	8.8	57.5	30.8	3.0
1 adult, 1+ child	48	31.2	27.1	37.5	4.2
2 adults, 0 child	1027	20.0	64.8	12.0	3.3
2 adults, 1+ child	929	22.3	56.8	19.1	1.8
3 adults, 0 child					
3 adults, 1+ child					
Total	3400	16.2	59.1	21.9	2.8

Note: Blank cells indicate suppression due to small or missing cell counts.

Table A.2: Population Count Comparison for Resolved True Positive AR Occupied Cases - Arizona Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	3400	5.2	11.0	59.1	14.6	7.3	2.8

Note: Table should be read as AR count relative to NRFU count.

Table A.3 illustrates the degree to which household composition agree in administrative records

and 2015 Census Test responses. Note that Table A.3 compares agreement in household composition, rather than counts conditional on composition detailed in Table A.1 and A.2. In particular, 2 adult and 0 children households agree in AR and NRFU in 57.3 percent of cases, which is comparable to a model estimated using all ACS respondents.

Table A.3: Household Composition Comparison AR vs. NRFU - Arizona Sample

	NRFU								
AR	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	48.4	2.7	15.7	4.4	2.1	0.7	16.3	1.0	8.8
1 adult, 1+ child	12.5	47.9	6.2	8.3	4.2	0.0	8.3	0.0	12.5
2 adults, 0 child	12.4	0.7	57.3	3.8	4.6	1.0	12.5	1.9	5.9
2 adults, 1+ child	3.7	5.9	3.3	62.1	0.6	5.0	14.0	1.8	3.6
3 adults, 0 child									
3 adults, 1+ child									
Total	24.6	3.6	24.7	20.0	2.7	2.0	14.4	1.5	6.5

Note: Blank cells indicate suppression due to small or missing cell counts.

A.2 ACS - April Only

Similarly, Tables A.4 to A.6 detail the ability of the model to correctly predict 2015 Census Test cases when ACS data is restricted to respondents that received notification to complete the survey by mail in April. Population counts by household composition are similar to those found in the Arizona analysis above. For example, conditional on a household composition of 2 adults and 1+ children, restricting model analysis to April data results in a match rate of 55.0 percent. In contrast, the Arizona-only model accurately captures count in 56.2 percent of cases. Tables A.5 and A.6 show that count and compositions match rates are similar to the ACS baseline case described Section 4.2.

Table A.4: Population Count Comparison by AR Household Composition- April Panel

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	1320	9.2	58.4	29.1	3.3
1 adult, 1+ child	77	24.7	31.2	40.3	3.9
2 adults, 0 child	1040	20.1	63.9	12.9	3.1
2 adults, 1+ child	958	24.0	55.0	19.0	2.0
3 adults, 0 child					
3 adults, 1+ child					
Total	3400	17.1	58.5	21.5	2.9

Note: Blank cells indicate suppression due to small or missing cell counts.

Table A.5: Population Count Comparison for Resolved True Positive AR Occupied Cases - April Panel

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	3400	5.8	11.3	58.5	14.8	6.7	2.9

Note: Table should be read as AR count relative to NRFU count.

Table A.6: Household Composition Comparison AR vs. NRFU - April Panel

AR	NRFU 1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	49.1	2.3	15.7	3.6	1.9	0.5	16.6	1.1	9.2
1 adult, 1+ child	6.5	41.6	5.2	14.3	2.6	3.9	13.0	2.6	10.4
2 adults, 0 child	12.4	0.7	56.2	4.1	5.1	1.0	12.7	1.7	6.1
2 adults, 1+ child	3.7	6.5	3.9	59.5	1.0	4.9	14.0	2.1	4.5
3 adults, 0 child									
3 adults, 1+ child									
Total	24.0	3.9	24.5	19.7	2.7	2.0	14.6	1.6	6.9

Note: Blank cells indicate suppression due to small or missing cell counts.

A.3 ACS - April Only Respondents

Table A.7: Population Count Comparison by AR Household Composition- April Panel (Responses only)

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	1406	9.2	57.1	30.5	3.1
1 adult, 1+ child	78	29.5	28.2	41.0	1.3
2 adults, 0 child	992	19.6	64.8	12.7	2.9
2 adults, 1+ child	922	23.5	55.9	18.5	2.1
3 adults, 0 child					
Total	3400	16.6	58.4	22.3	2.7

Note: Blank cells indicate suppression due to small or missing cell counts.

Lastly, we note that timing of NRFU differs substanially in the Decennial Census and in the ACS. In the previous section, we illustrated that restricting model estimation to a sample of ACS respondents that were sampled in April does not qualitatively change our findings. However, April ACS sampling results in NRFU followup in June – or two months after the intended reference period. Tables A.7 to A.9 detail agreement between administrative records and household response rates when the reference period is April. This scenario is more similar to the reference period of the decennial data collection.

Table A.8: Population Count Comparison for Resolved True Positive AR Occupied Cases - April Panel (Responses only)

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	3400	5.7	10.9	58.4	15.0	7.3	2.7

Note: Table should be read as AR count relative to NRFU count.

Tables A.7 and A.8 illustrate that the match rates are largely invariant to the two April sampling definitions with an overall match rate of approximately 58 percent.

A comparison of household composition allocations in this scenario shows a largely similar pattern to earlier scenarios presented in this Appendix. For example, 57.6 percent of 2 adult and 0 child cases in the administrative records agree with those found in NRFU. This is illustrated in Table A.9. In comparison, when ACS data was restricted to respondents sampled in April, the agreement rate for this household composition category is 56.2 percent.

Table A.9: Household Composition Comparison AR vs. NRFU - April Panel (Responses only)

AR	NRFU								
	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	47.8	2.7	15.6	4.5	2.1	0.6	16.6	1.0	9.2
1 adult, 1+ child	7.7	35.9	6.4	15.4	2.6	2.6	14.1	3.8	11.5
2 adults, 0 child	11.9	0.6	57.6	4.2	4.9	0.9	12.2	1.7	5.9
2 adults, 1+ child	3.5	6.0	3.8	60.7	0.9	4.8	13.8	1.8	4.8
3 adults, 0 child									
Total	24.4	3.7	24.4	19.9	2.6	1.9	14.5	1.5	7.1

Note: Blank cells indicate suppression due to small or missing cell counts.

B. Appendix: 2016 Census Test Extension - Los Angeles

B.1 2016 Census Test Extensions: Methodology

As an extension to the work done using the 2015 Census Test, a similar analysis was performed using the 2014 ACS and the 2016 Census Test. The 2016 Census Test differed in geographic location, focusing on Los Angeles and Houston as test cities. Because of the geographic locations, the sample sizes for both 2016 Census tests are much larger. Each of the samples for Los Angeles and Houston were roughly three times the size of the sample in the 2015 test, which covered Maricopa County in Arizona. The cutoff point for the distance function is different from the 2015 test, reflecting the need to remove a similar percentage of the NRFU workload. For Los Angeles and Houston, the cutoff is 6,200 to correspond with removal of about 10 percent of the NRFU workload, while previously for Arizona and the 2015 test, the corresponding cutoff was 3400.

There is also a major difference in the sampling frame for the NRFU operation which reflects major changes in the contact strategies employed in 2016. Using the 2010 Census-trained models developed and tested in 2015, 6,200 NRFU addresses were flagged as occupied and enumerated using AR instead of receiving NRFU interview attempts. In order to assess the performance of the field-removal using AR, a 1 in 5 systematic sample (by occupied distance) of these 6,200 addresses still received full NRFU contact attempts to facilitate comparison with AR enumeration. All other NRFU addresses not included in the sample of 6,200 received full NRFU contact attempts.

The lack of NRFU responses for 80 percent of the 6,200 AR-occupied addresses complicates the comparison of an ACS removal list with the 2010 Census removal list. A household which was selected by the original 2010 Census model has a 20 percent chance of receiving the full NRFU contact strategy which allows us to evaluate whether AR count and composition matched the reported count and composition in the Census Test. A household which is selected for removal by ACS-trained models may or may not have been selected by the 2010 Census model for removal.

If the ACS-removed household was not in the original 2010 Census removal list, it received the full NRFU contact strategy with certainty and can be evaluated. If the ACS-removed household was also part of the 2010 Census removal list, it can only be evaluated against the NRFU response if it was one of the 20 percent of addresses receiving NRFU contacts. The ACS removal list is thus composed of addresses which received NRFU contact attempts with certainty as well as some addresses which only received NRFU contact with 20 percent probability. To adjust for the differences in probability that these addresses in the Census-removed list had NRFU contacts, any 2010 Census-removed address with a NRFU contact is weighted up by a factor of 5 to facilitate comparison between the two cases. (The 2010 Census-removed addresses with no NRFU contact are dropped, effectively giving them a weight of zero.)

B.2 2016 Census Test Extension - Los Angeles

In this section, we replicate our main results using the 2016 Census Test for Los Angeles. When the results of the 2016 Census Test for Los Angeles are examined and compared to the 2015 Census Test, some differences emerge. Based on the result shown in Table B.1, for the Household-Composition model, the 2016 test has a lower match rate than the 2015 test. However, the opposite is true for the Person-Place model. The 2016 LA test performs with a higher match rate compared to the 2015 test.

Table B.1: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	46.9	61.0
Baseline - ACS	45.9	59.9

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4216 (ACS) and 5518 (Census)

Table B.2 shows the comparison of population counts between AR and NRFU. This table is analogous to Table 1 for the 2015 analysis. Compared to the 2015 results, there is a higher match rate between AR enumeration and NRFU counts of addresses in Los Angeles in 2016. This is true among all household composition categories. The percentage of addresses that are equal is consistently between 60 and 65 percent across all household composition categories. This result differs from the 2015 analysis, where there tended to be an overcount for larger AR households.

Table B.3 displays the population count comparison for resolved true positive AR occupied cases. Overall, there is a 61 percent match rate between AR and NRFU counts. The results show that 3.7 percent of AR overcount household population by 2 or more individuals and 6.5 percent of AR overcount household population by one individual. Undercounts are more prevalent in the results, with a 12.7 percent AR undercount of household population by one individual and a 5.7 percent AR undercount of 2 or more individuals. The percent of NRFU count unknown is 10.4 percent, which is higher than the percentage in the 2015 analysis.

Table B.2: Population Count Comparison by AR Household Composition- Full Census 2010 Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	2800	3.7	60.8	22.9	12.7
1 adult, 1+ child					
2 adults, 0 child	1096	13.7	60.1	16.1	10.1
2 adults, 1+ child	2247	16.7	61.5	14.0	7.7
3 adults, 1+ child					
Total	6157	10.2	61.0	18.4	10.4

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.

Number of units employed: 5518

Blank cells indicate suppression due to small or missing cell counts.

Table B.3: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full Census 2010 Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	6157	3.7	6.5	61.0	12.7	5.7	10.4

Note: Table should be read as AR count relative to NRFU count.

Table B.4 compares the household composition between AR and NRFU for the full Census 2010 sample. The rows correspond to household composition categories for AR and the columns represent the categories for NRFU. Dominant diagonal values represent a high match rate between AR and NRFU across household composition categories. It is important to note that the "Unknown age" category in NRFU is higher for the 2016 Los Angeles test than in the 2015 test (roughly 37 percent in 2016 versus 14 percent in 2015).

Table B.4: Household Composition Comparison AR vs. NRFU - Full Census 2010 Sample

	NRFU								
AR	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	46.1	0.2	9.8	0.6	0.7	0.3	37.6	1.0	3.7
1 adult, 1+ child									
2 adults, 0 child	6.5	0.0	46.1	1.6	4.6	1.9	37.0	0.7	1.6
2 adults, 1+ child	2.1	3.3	1.3	48.1	0.0	3.8	37.3	1.4	2.6
3 adults, 1+ child									
Total	22.9	1.5	13.1	18.1	1.1	1.9	37.4	1.1	2.9

Note: Blank cells indicate suppression due to small or missing cell counts.

Similar tables are produced for the ACS sample, with Table B.5 showing the population count comparison by AR household composition. This table shows that there are only three categories of household composition in the ACS sample: 1 adult, 0 children; 2 adults, 0 children; and 2 adults, 1 or more children. For the entire sample, the match rate between AR and NRFU is about 60 percent. This is similar to the result for the Census sample, which is shown in Table B.2. Table B.6 indicates that there is a higher percentage of undercounting in the AR for the ACS sample. The dominant diagonal in Table B.7 shows a high match rate between AR and NRFU with those values ranging from 43.1 to 50.1 percent. Again, there is roughly a third of the sample in the "Unknown age" category.

Similar to Tables 8 and 9, Tables B.8 and B.9 show the overlap in the workload removal from the 2010 Census and ACS trained models, respectively. Each table ranks the 6,200 records removed from the workload by ascending deciles of the distance function value, with the first decile being the records removed with the greatest degree of confidence. For both the Census and ACS trained models, the overlap is close to 100 percent in the first decile and roughly 50 percent in the tenth decile. The ACS cut-off distances are much lower than the Census cut-off distances.

Table B.10 compares the match rates for counts and compositions for different training samples of the 2014 ACS. The baseline case in the above analysis is similar to the match rates attained by training on refined subsamples of the 2014 ACS, including restricting the training data by state or month-in-sample.

Table B.5: Population Count Comparison by AR Household Composition- Full ACS Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	2765	3.9	59.0	24.7	12.5
2 adults, 0 child	1577	17.6	56.8	15.3	10.3
2 adults, 1+ child	1777	13.6	64.0	14.4	8.1
Total	6119	10.2	59.9	19.3	10.7

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4216

Table B.6: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full ACS Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	6119	2.4	7.8	59.9	13.4	5.9	10.7

Note: Table should be read as AR count relative to NRFU count.

Table B.7: Household Composition Comparison AR vs. NRFU - Full ACS Sample

AR	NRFU								
	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	44.8	0.4	10.4	1.0	1.1	0.5	37.0	0.9	3.9
2 adults, 0 child	9.5	0.3	43.1	1.0	5.3	1.2	37.0	0.6	2.0
2 adults, 1+ child	2.0	3.5	1.0	50.1	0.0	4.5	36.2	1.5	1.2
Total	23.3	1.3	16.1	15.2	1.9	1.8	36.8	1.0	2.6

Table B.8: Sample Overlap - Census vs. ACS (by Census distance)

Decile of Census Distance	% present in ACS sample	Cut-off Census distances
1 - Highest precision	99.2	0.417
2	97.6	0.461
3	86.2	0.496
4	75.6	0.531
5	79.7	0.552
6	67.5	0.576
7	66.7	0.592
8	56.5	0.608
9	48.6	0.625
10 - Lowest precision	45.5	0.640

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4216 (ACS) and 5518 (Census)

Table B.9: Sample Overlap - Census vs. ACS (by ACS distance)

Decile of ACS Distance	% present in Census sample	Cut-off ACS distances
1 - Highest precision	99.0	0.221
2	91.4	0.250
3	90.8	0.276
4	69.8	0.300
5	87.4	0.315
6	74.0	0.336
7	62.4	0.355
8	56.0	0.375
9	47.1	0.395
10 - Lowest precision	49.9	0.416

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4216 (ACS) and 6207 (Census)

Table B.10: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	46.9	61.0
Baseline - ACS	45.9	59.9
ACS - State-specific	46.7	60.9
ACS - April panel (responses)	46.1	59.9
ACS - February panel	46.7	59.9
ACS - March panel	46.6	60.1
ACS - April panel	46.8	60.1
ACS - May panel	46.3	59.5
ACS - June panel	46.5	60.0
ACS - July panel	46.3	59.9
ACS - NRFU	44.6	58.2

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4216 (ACS) and 5518 (Census).
NRFU training based on ACS mafids where telephone and in-person follow-ups
(CATI and CAPI) were conducted due to non-response of the mail-in form.

The next several figures are used to investigate the sensitivity of the matching results to the weighting of the Person-Place and Household-Composition models in the distance function. As in the 2015 analysis, an equal weighting of the predicted probabilities for each model is assumed in the 2016 analysis. For the sensitivity analysis, the full range of alternate weights are considered. For the 2010 Census and ACS trained models, the Person-Place and Household-Composition model match rates are calculated for each scenario and then graphed by δ in Figures 7, 8, 9, and 10. Overall, the shape for all these figures is an inverted "U" shape. The graphs for the ACS model are flatter in the middle compared to the ones for the Census model. However, all the figures have a peak to the left of center, which indicates that instead of giving equal weighting to the Person-Place and Household-Composition models, more weight should be placed on the Person-Place model. This differs from the 2015 results, where equal weighting was preferred.

Figure 7: Household Composition Match Rates and Distance Weights (ACS Model)

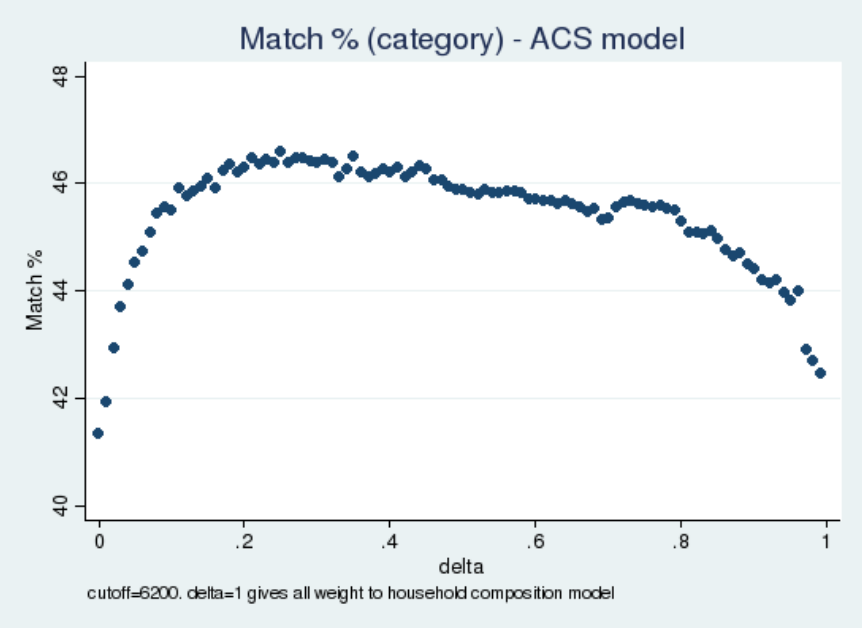


Figure 8: Household Unit Count Match Rates and Distance Weights (ACS Model)

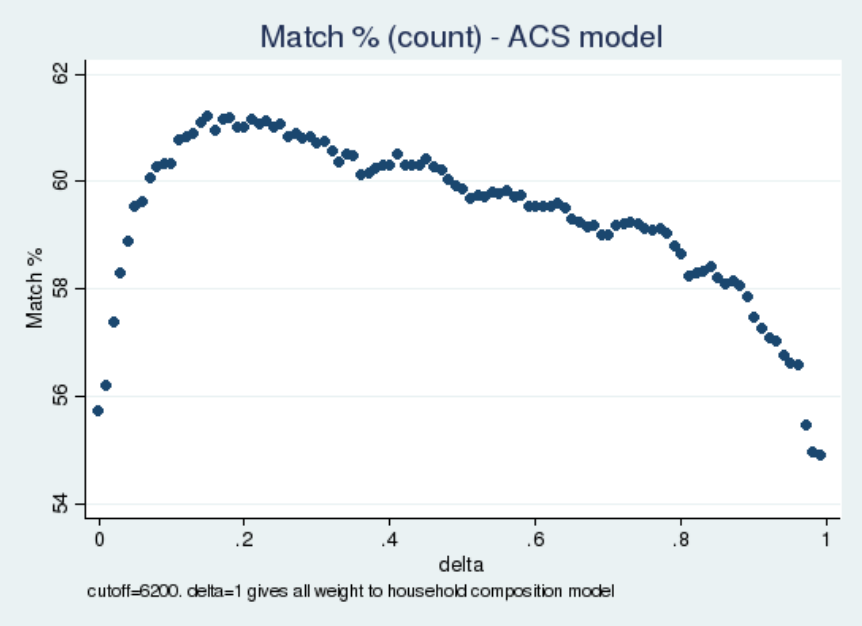


Figure 9: Household Composition Match Rates and Distance Weights (Census Model)

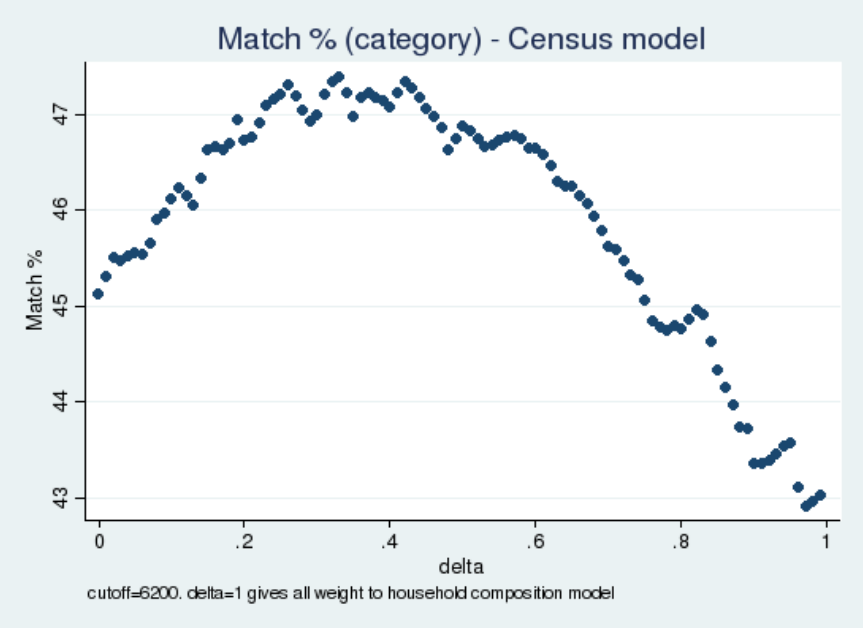
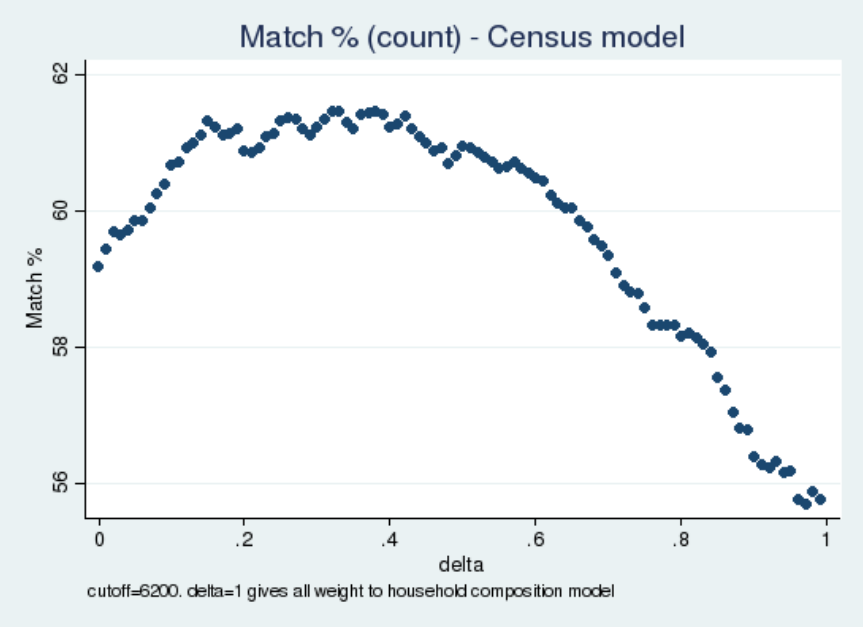


Figure 10: Household Unit Count Match Rates and Distance Weights (Census Model)



C. Appendix: 2016 Census Test Extension - Houston

In this section, we replicate all our main results using the 2016 Census Test for Houston. If we compare the results for Houston in Table C.1 to the results of the Los Angeles test in Table B.1 and also to the results for the 2015 test, the match rates are lowest for the Houston test.

Table C.1: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	42.6	54.2
Baseline - ACS	44.5	55.3

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 3420 (ACS) and 4892 (Census)

When taking a closer look at the comparison of population counts between AR and NRFU, there are only three household composition categories: 1 adult, 0 children; 2 adults, 0 children; and 2 adults, 1 or more children. The counts are equal roughly 50 to 60 percent of the time, which is slightly lower than what we saw in Los Angeles. Overall, there is a match rate of 54.2 percent for population count in Houston.

Table C.2: Population Count Comparison by AR Household Composition- Full Census 2010 Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	4824	3.5	52.4	16.3	27.7
2 adults, 0 child	369	14.6	62.3	7.6	15.4
2 adults, 1+ child	990	19.3	59.8	8.6	12.3
Total	6183	6.7	54.2	14.6	24.5

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4892

Table C.3 shows the breakdown in population count differences, and similar to the Los Angeles results, there is a larger amount of undercounting than overcounting in the AR. The percent of NRFU count unknown is 24.5 percent, which is higher than both the percentage in the Los Angeles test and in the 2015 test. Table C.4 demonstrates a dominant diagonal, similar to the results in the analogous tables for the Los Angeles and 2015 tests. Among the cells that lie on the diagonal, the match rate for the 1 adult, 0 children category is lowest at 39.4 percent. For the other two categories, the match rate is higher and more in line with the results from Los Angeles. The match rate for 2 adults, 0 children is 50.7 percent and for 2 adults, 1 or more children it is 54.9 percent.

Table C.3: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full Census 2010 Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	6183	2.2	4.5	54.2	11.8	2.8	24.5

Note: Table should be read as AR count relative to NRFU count.

Table C.4: Household Composition Comparison AR vs. NRFU - Full Census 2010 Sample

AR	NRFU								
	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	39.4	0.4	8.9	1.0	0.4	0.1	46.0	0.2	3.5
2 adults, 0 child	8.7	0.3	50.7	2.2	1.1	0.0	36.9	0.0	0.3
2 adults, 1+ child	1.7	2.2	2.1	54.9	0.0	2.9	33.3	0.0	2.7
Total	31.6	0.7	10.3	9.7	0.4	0.6	43.4	0.2	3.2

For the population count comparison with the ACS sample in Table C.5, an additional household composition category is included in the AR occupied removal list: 1 adult, 1 or more children. However, figures for this category are not shown in the table because they account for less than 1 percent of the evaluation sample. Overall, the match rate is 55.3 percent, which is similar to and slightly higher than the match rate for the Census sample. There is a larger amount of undercounting than overcounting in AR for those households with one adult, while the reverse is true for households with two adults. The percent of NRFU count unknown is similar to the percentage for the Census results in Houston.

Looking at Table C.7, once again, the household composition comparison shows a dominant diagonal for all categories, with cell values ranging from 41 to 55 percent.

Tables C.8 and C.9 show the overlap in the workload removal from the 2010 Census and ACS trained models, respectively. For both the Census and ACS trained models, the overlap is close to 98 percent in the first decile and roughly 40 to 50 percent in the tenth decile. The ACS cut-off distances are much lower than the Census cut-off distances, which is consistent with the Los Angeles results.

Table C.10 compares the match rates for counts and compositions for different training samples of the 2014 ACS. The baseline case in the above analysis is similar to the match rates attained by training on refined subsamples of the 2014 ACS, including restricting the training data by state or month-in-sample.

Table C.5: Population Count Comparison by AR Household Composition- Full ACS Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	3963	4.1	54.2	15.2	26.5
1 adult, 1+ child					
2 adults, 0 child	1090	17.2	57.1	6.9	18.8
2 adults, 1+ child	1144	21.0	57.6	8.2	13.2
Total	6200	9.5	55.3	12.4	22.7

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.

Number of units employed: 3420

Blank cells indicate suppression due to small or missing cell counts.

Table C.6: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full ACS Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	6200	2.7	6.8	55.3	10.0	2.4	22.7

Note: Table should be read as AR count relative to NRFU count.

Table C.7: Household Composition Comparison AR vs. NRFU - Full ACS Sample

AR	NRFU 1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	41.0	0.4	7.9	1.1	0.3	0.2	44.8	0.2	4.1
1 adult, 1+ child									
2 adults, 0 child	8.6	0.3	46.8	1.5	2.2	0.3	38.4	0.0	1.9
2 adults, 1+ child	1.8	2.4	1.7	55.0	0.1	2.5	32.9	0.3	3.3
Total	28.0	0.8	13.6	11.1	0.6	0.6	41.5	0.2	3.6

Note: Blank cells indicate suppression due to small or missing cell counts.

Table C.8: Sample Overlap - Census vs. ACS (by Census distance)

Decile of Census Distance	% present in ACS sample	Cut-off Census distances
1 - Highest precision	97.6	0.386
2	85.5	0.448
3	80.5	0.481
4	75.0	0.504
5	66.1	0.530
6	57.0	0.554
7	52.2	0.580
8	49.8	0.606
9	51.8	0.631
10 - Lowest precision	49.7	0.654

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 3420 (ACS) and 4892 (Census)

Table C.9: Sample Overlap - Census vs. ACS (by ACS distance)

Decile of ACS Distance	% present in Census sample	Cut-off ACS distances
1 - Highest precision	97.4	0.209
2	87.4	0.239
3	80.8	0.264
4	50.6	0.292
5	60.9	0.309
6	60.0	0.326
7	63.9	0.343
8	61.0	0.357
9	60.2	0.375
10 - Lowest precision	41.3	0.390

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 3420 (ACS) and 5484 (Census)

Table C.10: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	42.6	54.2
Baseline - ACS	44.5	55.3
ACS - State-specific	44.6	55.7
ACS - April panel (responses)	43.8	54.3
ACS - February panel	44.6	55.4
ACS - March panel	44.2	54.7
ACS - April panel	44.6	55.5
ACS - May panel	44.5	55.3
ACS - June panel	44.5	55.4
ACS - July panel	44.5	55.4
ACS - NRFU	42.9	53.6

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 3420 (ACS) and 4892 (Census).
NRFU training based on ACS mafids where telephone and in-person follow-ups
(CATI and CAPI) were conducted due to non-response of the mail-in form.

Figure 11: Household Composition Match Rates and Distance Weights (ACS Model)

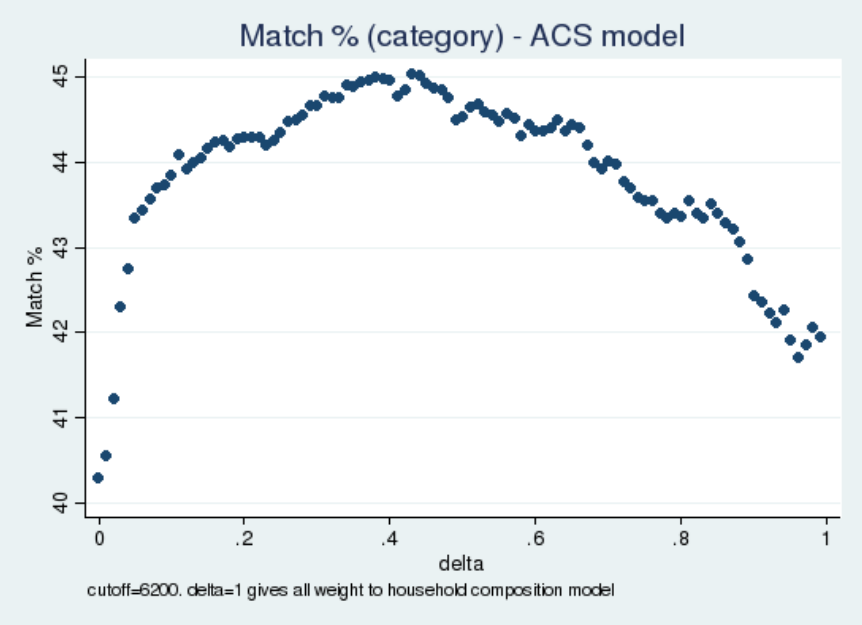


Figure 12: Household Unit Count Match Rates and Distance Weights (ACS Model)

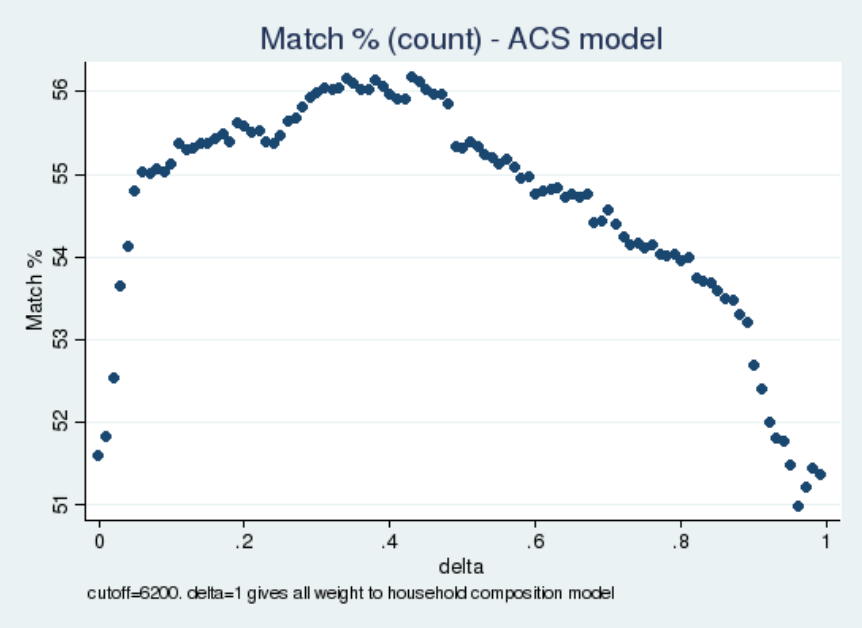


Figure 13: Household Composition Match Rates and Distance Weights (Census Model)

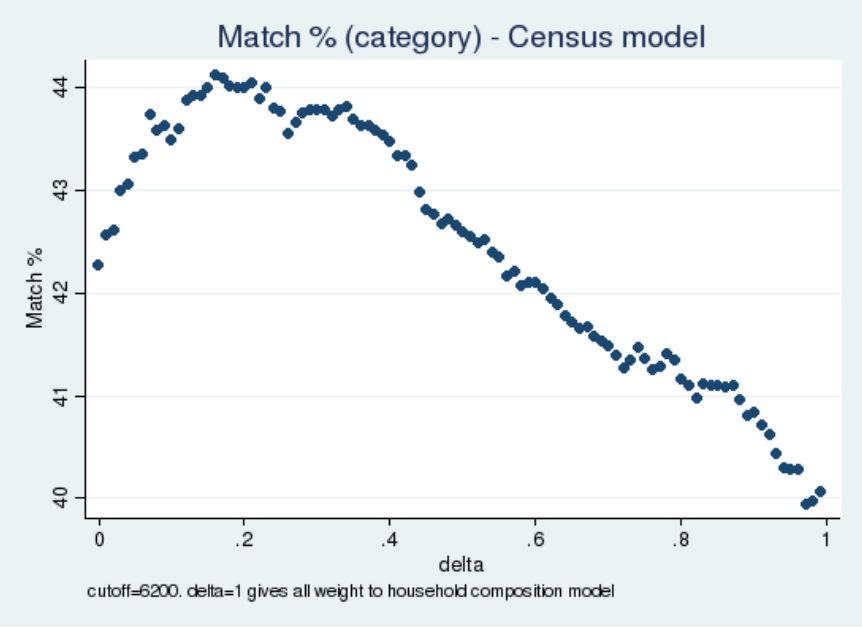
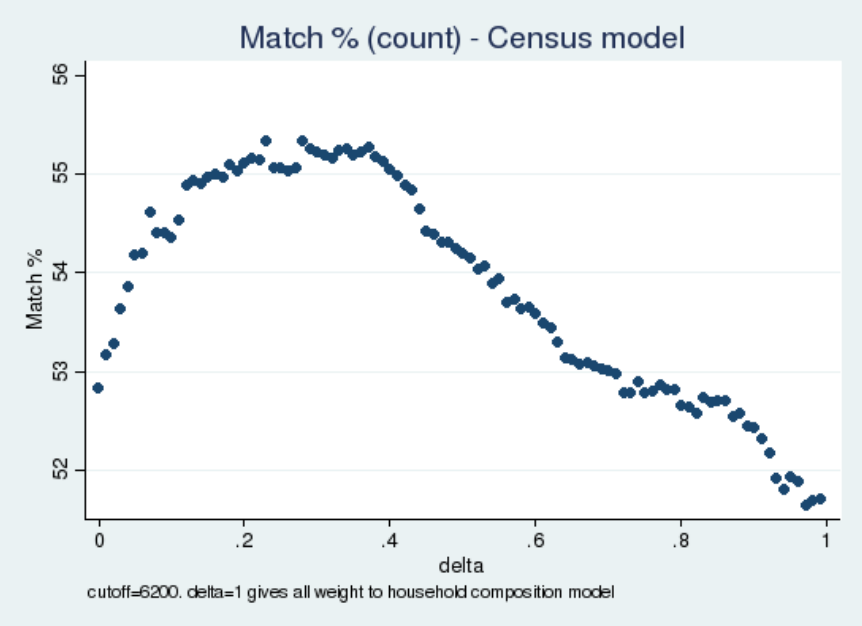


Figure 14: Household Unit Count Match Rates and Distance Weights (Census Model)



Figures 11, 12, 13, and 14 show the distance function sensitivity analysis for the Houston test. As in the 2015 test and the 2016 Los Angeles test, the shape for all these figures is an inverted "U" shape. The graphs for the ACS model peak when δ is around 0.4. The graphs for the ACS model are also less flat than the ones for the Los Angeles test. The Census model shows a peak when delta is roughly 0.2 for the category match percentage and a peak delta around 0.3 for the count match percentage. All the figures have a peak to the left of center, which indicates that instead of giving equal weighting to the Person-Place and Household-Composition models, more weight should be placed on the Person-Place model. This differs from the 2015 results, where equal weighting was preferred, but is consistent with the Los Angeles test.

D. Appendix: 2014 ACS Extension SNAP analysis

One advantage of using ACS data in lieu of the 2010 Decennial Census for model training is the ability to include variables from data sources which may not be available back to 2010. One example of such data is state-provided records of enrollment in Supplemental Nutrition Assistance Programs (SNAP). While Census has been provided with a range of SNAP data from several states, few states have made data available on SNAP coverage extending back to 2010. Being able to use more recent vintages of ACS data would increase the usability of data which may only be available with limited time-series coverage. In this exercise, we investigate the use of ACS in training a model which includes SNAP data from Illinois.

D.1 Methodology

To test the use of ACS data merged with SNAP variables from Illinois requires substantial changes to the methodology which we used in prior studies. The main limitation which prevents replicating our 2015 or 2016 Census test results is the fact that SNAP data was not available for any of the state-year combinations of the 2015 (Arizona) or 2016 (California, Texas) Census tests. Since our SNAP data set is for Illinois in 2014, we instead use ACS data for both training and evaluation. We take all ACS observations in Illinois in 2014 and randomly subset addresses into a training sample and an evaluation sample, each containing 50 percent of all records.[11] We fit the PP and HC model on the training subset. To evaluate, we score the fitted model on the evaluation subset of ACS observations and calculate the Euclidian distance to determine candidates for AR occupied removal. We remove 6400 addresses with the smallest distances and compare the AR counts and compositions of these addresses to the household counts and compositions reported in the ACS data.[12]

To test the possible improvement of the model by including SNAP data, we fit the model on ACS and AR data with and without the inclusion of SNAP variables.[13] We score the evaluation subset of ACS data using both models and flag the 6400 lowest-distance (highest certainty) addresses for AR occupied removal and enumeration. We compare the households which are removed from the evaluation subsample in both cases by comparing the agreement rate of the AR counts and compositions with the actual counts and compositions reported in the ACS. We find that inclusion of SNAP variables in the training of the AR occupied models does not improve the match rate relative to the ACS-trained models with no SNAP variables.

The model's lack of improvement from inclusion of SNAP variables is due to the fact that high certainty addresses which are selected for AR occupied removal have a low percentage of SNAP

[11] Each sample contains 33,206 records.

[12] Note that we do not remove any addresses using the vacancy model as was done in our 2015 and 2016 Census test cases.

[13] For the PP model, we employ two SNAP variables: a person-level SNAP recipient indicator and an indicator for whether a PIK could be found as a SNAP recipient in a different MAFID. For the HC model, we employ two analogous household-level SNAP variables: an indicator of whether anyone in a household was a SNAP recipient and an indicator for whether a PIK in a unit could be found in the SNAP files at a different MAFID.

records. While SNAP data may improve the accuracy of estimates in determining AR concordance of some occupied households, it does so primarily for households which have high distance values which are not selected for removal.

D.2 Results

Table D.1 displays the category and count match rates for our two cases, training and evaluating the model on two subsets of ACS observations. We compare the marginal change due to incorporating SNAP data by comparing our baseline model with no SNAP variables (NO SNAP) and including SNAP variables in our model covariates (SNAP). The match rates for both count and composition remain nearly unchanged between the two cases.

Table D.1: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
ACS 2014 NO SNAP	92.0	90.2
ACS 2014 SNAP	92.0	90.1

Note: Distance cutoff set at 6400 units. Observations weighted based on Census Test removal.
Number of units employed: 6400 (ACS) and (Census)

Looking at the composition and counts of the 6400 removed records in the SNAP and No SNAP cases, the differences are negligible. A comparison between Tables D.2-D.4 (No SNAP case) and Tables D.5-D.7 (SNAP case) shows that both cases have similar distributions of household compositions removed based on administrative records. In both cases, no households with AR compositions of 1 adult and 1 or more child were selected for removal.

Table D.2: Population Count Comparison by AR Household Composition- ACS 2014 training without SNAP

Household Composition	Units	Greater in AR	Equal	Fewer in AR	ACS Count Unknown
	N	%	%	%	%
1 adult, 0 child	3106	2.6	91.3	6.1	0.0
2 adults, 0 child	1194	5.7	91.7	2.6	0.0
2 adults, 1+ child	2100	6.5	87.7	5.8	0.0
Total	6400	4.5	90.2	5.4	0.0

One reason for such a small marginal change in the removal samples between SNAP and No SNAP cases is the lack of SNAP enrollees in high precision (low distance) addresses. Table D.8 lists deciles by distance of the removed subset for the SNAP case. Accross all distance deciles of the removal subsample, the percentage of addresses with SNAP observations present is only around 2 percent. This compares to the overall percentage of SNAP observations in all addresses of about 6 percent.

Table D.3: Population Count Comparison for Resolved True Positive AR Occupied Cases - ACS 2014 training without SNAP

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	ACS Count Unknown
	N	%	%	%	%	%	%
Control Panel	6400	1.8	2.7	90.2	4.7	0.7	0.0

Note: Table should be read as AR count relative to ACS count.

Table D.4: Household Composition Comparison AR vs. ACS - ACS 2014 training without SNAP

AR	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	ACS codes 0 & 10
	ACS						
1 adult, 0 child	91.2	0.1	5.5	0.1	0.3	0.0	2.7
2 adults, 0 child	4.0	0.0	91.6	0.3	2.1	0.0	2.0
2 adults, 1+ child	0.5	1.3	0.5	93.2	0.1	0.5	3.8
Total	45.2	0.5	19.9	30.7	0.6	0.2	2.9

Table D.5: Population Count Comparison by AR Household Composition- ACS 2014 training with SNAP

Household Composition	Units	Greater in AR	Equal	Fewer in AR	ACS Count Unknown
	N	%	%	%	%
1 adult, 0 child	3090	2.6	91.3	6.1	0.0
2 adults, 0 child	1253	5.5	91.8	2.7	0.0
2 adults, 1+ child	2057	6.4	87.4	6.2	0.0
Total	6400	4.4	90.1	5.5	0.0

Table D.6: Population Count Comparison for Resolved True Positive AR Occupied Cases - ACS 2014 training with SNAP

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	ACS Count Unknown
	N	%	%	%	%	%	%
Control Panel	6400	1.7	2.7	90.1	4.8	0.7	0.0

Note: Table should be read as AR count relative to ACS count.

Table D.7: Household Composition Comparison AR vs. ACS - ACS 2014 training with SNAP

AR	ACS 1 adult, 0 child	ACS 1 adult, 1+ child	ACS 2 adults, 0 child	ACS 2 adults, 1+ child	ACS 3 adults, 0 child	ACS 3 adults, 1+ child	ACS codes 0 & 10
1 adult, 0 child	91.2	0.1	5.6	0.1	0.2	0.0	2.8
2 adults, 0 child	3.8	0.0	91.7	0.2	2.2	0.0	2.1
2 adults, 1+ child	0.5	1.3	0.5	93.3	0.1	0.6	3.6
Total	45.0	0.5	20.8	30.1	0.6	0.2	2.9

Table D.9 lists addresses by distance deciles for the entire sample. The top 2 deciles in terms of certainty (lowest distances) comprise the removal sample (see Table D.8), and have a low percentage of SNAP observations. The percentage of addresses with SNAP observations is much higher for the lower certainty, high distance deciles. SNAP variables are thus more likely to be relevant and present in the portion of the ACS sample which was not selected for AR occupied removal. Table D.10 shows that this is also the case for the No SNAP training case. Comparing the distances at each decile between Table D.9 and Table D.10, the calculated distances show very little change by adding SNAP variables in our regression specifications. Including SNAP doesn't seem to meaningfully improve the certainty of AR occupied cases even at larger distances.

Table D.8: SNAP % by ACS Distance (w/ SNAP in training) - Removed Sample Only

Decile of ACS Distance	% SNAP	Cut-off ACS distances
1 - Highest precision	1.09	0.150
2	2.03	0.169
3	0.62	0.178
4	1.09	0.182
5	1.56	0.192
6	0.47	0.206
7	2.34	0.216
8	1.88	0.223
9	1.25	0.232
10 - Lowest precision	1.72	0.237

D.3 Using a "NRFU" sample for evaluation

The count and composition match rates which result from training and evaluating on ACS are much higher than those found in the Census Test cases. Although our interest is in the marginal

Table D.9: SNAP % by ACS Distance (w/ SNAP in training) - Full Sample

Decile of ACS Distance	% SNAP	Cut-off ACS distances
1 - Highest precision	1.18	0.191
2	1.69	0.236
3	0.57	0.249
4	2.07	0.299
5	4.36	0.375
6	6.30	0.516
7	6.53	0.688
8	6.75	0.851
9	7.48	1.042
10 - Lowest precision	5.73	1.385

Table D.10: SNAP % by ACS Distance (w/o SNAP in training) - Full Sample

Decile of ACS Distance	% SNAP	Cut-off ACS distances
1 - Highest precision	1.56	0.191
2	1.94	0.237
3	1.21	0.249
4	3.12	0.299
5	3.76	0.376
6	5.82	0.517
7	5.95	0.687
8	6.46	0.850
9	7.26	1.041
10 - Lowest precision	5.57	1.385

change between the NO SNAP and SNAP cases rather than the level, we address the concern that the ACS subsample differs significantly from the NRFU samples in Census Tests. The differences in sampling and contact methodology between ACS and Census Tests could result in different overall match rates and potentially different marginal improvements from adding SNAP variables. To address this concern, we train our model using the same ACS training subset, but change our evaluation subset to only include addresses which were contacted in "CAPI" and "CATI," the telephone and in-person follow-up processes in the ACS. This sample should more closely resemble the NRFU samples which form the evaluation data for the 2015 and 2016 Census tests.

In Table D.11, the match rate resulting from the "NRFU" subsample exhibits a much lower overall match rate. The marginal change to the match rate by including SNAP variables remains insignificant. The match rate for the NRFU evaluation sample is much lower because as a whole, the NRFU evaluation sample has a much lower match rate (before removal). The household composition match rate for the full ACS sample is 58 percent, while the analogous match rate for the ACS-NRFU subsample is 31 percent (pre-removal figures not shown in Table D.11).

Table D.11: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
ACS 2014 NO SNAP	92.0	90.2
ACS 2014 SNAP	92.0	90.1
ACS 2014 (NRFU ONLY) NO SNAP	31.3	31.2
ACS 2014 (NRFU ONLY) SNAP	31.3	31.2

Note: Distance cutoff set at 6400 units. Observations weighted based on Census Test removal.
Number of units employed: 6400 (ACS) and (Census)

E. Appendix: Evaluating ACS 2015 using the 2016 Census Test - Los Angeles

In this section, we detail the effects of using the 2015 ACS for training and evaluation on the 2016 Census Test in Los Angeles. We interpret the results presented in this section as not showing economically meaningful differences with those that were estimated using 2014 ACS and discussed in Appendix B.

Table E.1 shows that the overall match rates in counts and composition are slightly lower when estimated using the 2015 ACS than the 2010 Census. In comparison, comparing Table E.1 to Table B.1 shows that the differences between 2014 and 2015 ACS years are small. The remaining tables detail count and composition comparisons to 2010 Census training and do not illustrate any meaningful differences with those estimated using 2014 ACS.

Table E.1: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	46.9	61.0
Baseline - ACS 2015	46.5	60.7

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4318 (ACS) and 5518 (Census)

Table E.2: Population Count Comparison by AR Household Composition- Full Census 2010 Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	2800	3.7	60.8	22.9	12.7
1 adult, 1+ child					
2 adults, 0 child	1096	13.7	60.1	16.1	10.1
2 adults, 1+ child	2247	16.7	61.5	14.0	7.7
3 adults, 1+ child					
Total	6157	10.2	61.0	18.4	10.4

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 5518
Blank cells indicate suppression due to small or missing cell counts.

Table E.3: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full Census 2010 Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	6157	3.7	6.5	61.0	12.7	5.7	10.4

Note: Table should be read as AR count relative to NRFU count.

Table E.4: Household Composition Comparison AR vs. NRFU - Full Census 2010 Sample

AR	NRFU								
	1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	46.1	0.2	9.8	0.6	0.7	0.3	37.6	1.0	3.7
1 adult, 1+ child									
2 adults, 0 child	6.5	0.0	46.1	1.6	4.6	1.9	37.0	0.7	1.6
2 adults, 1+ child	2.1	3.3	1.3	48.1	0.0	3.8	37.3	1.4	2.6
3 adults, 1+ child									
Total	22.9	1.5	13.1	18.1	1.1	1.9	37.4	1.1	2.9

Note: Blank cells indicate suppression due to small or missing cell counts.

F. Appendix: ACS 2015 - 2016 Census Test Extension - Houston

In this section, we detail the effects of using the 2015 ACS for training and evaluation on the 2016 Census Test in Houston. We interpret the results presented in this section as not showing economically meaningful differences with those that were estimated using 2014 ACS and discussed in Appendix C.

Table F.1 shows that the overall match rates in counts are slightly lower when estimated using the 2015 ACS than the 2010 Census, but slightly ligher when comparing match rates in compositions. In comparison, comparing Table F.1 to Table C.1 shows that the differences between 2014 and 2015 ACS years are small. The remaining tables detail count and composition comparisons to 2010 Census training and do not illustrate any meaningful differences with those estimated using 2014 ACS.

Table F.1: Comparison of Matches in Household Composition

Training Data	Match % (category)	Match % (count)
Baseline - Census 2010	42.6	54.2
Baseline - ACS 2015	44.4	56.1

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 3507 (ACS) and 4892 (Census)

Table F.2: Population Count Comparison by AR Household Composition- Full Census 2010 Sample

Household Composition	Units	Greater in AR	Equal	Fewer in AR	NRFU Count Unknown
	N	%	%	%	%
1 adult, 0 child	4824	3.5	52.4	16.3	27.7
2 adults, 0 child	369	14.6	62.3	7.6	15.4
2 adults, 1+ child	990	19.3	59.8	8.6	12.3
Total	6183	6.7	54.2	14.6	24.5

Note: Distance cutoff set at 6200 units. Observations weighted based on Census Test removal.
Number of units employed: 4892

Table F.3: Population Count Comparison for Resolved True Positive AR Occupied Cases - Full Census 2010 Sample

	Units	2+ Greater in AR	1 Greater in AR	Match	1 Fewer in AR	2+ Fewer in AR	NRFU Count Unknown
	N	%	%	%	%	%	%
Control Panel	6183	2.2	4.5	54.2	11.8	2.8	24.5

Note: Table should be read as AR count relative to NRFU count.

Table F.4: Household Composition Comparison AR vs. NRFU - Full Census 2010 Sample

AR	NRFU 1 adult, 0 child	1 adult, 1+ child	2 adults, 0 child	2 adults, 1+ child	3 adults, 0 child	3 adults, 1+ child	Unknown age	Other	Not occupied
1 adult, 0 child	39.4	0.4	8.9	1.0	0.4	0.1	46.0	0.2	3.5
2 adults, 0 child	8.7	0.3	50.7	2.2	1.1	0.0	36.9	0.0	0.3
2 adults, 1+ child	1.7	2.2	2.1	54.9	0.0	2.9	33.3	0.0	2.7
Total	31.6	0.7	10.3	9.7	0.4	0.6	43.4	0.2	3.2