Explaining Differences Between Administrative Records Modeling Simulations and 2010 Census Results

Andrew Keller[1]
U.S. Census Bureau, 4600 Silver Hill Rd., Washington, DC 20233

Abstract

As part of the 2010 decennial census, about one third of the US population was enumerated by a personal visit during the Nonresponse Followup operation. A major driver of cost in the 2010 Census was the Nonresponse Followup operation with a total cost of about $1.6 billion. Consequently, the U.S. Census Bureau is researching the possible use of administrative records to provide an occupancy status and count for some Nonresponse Followup addresses. Since some of the households included in the Nonresponse Followup operation were present in administrative records, the Census Bureau is planning to enumerate some nonresponding households with administrative records. However, occasionally census and administrative records outcomes are different.

This paper seeks to understand why inconsistencies between census and administrative records data occur. This will help inform when administrative records should be used to enumerate the household as opposed to enumerating the household with an in-person followup interview. Specifically, the analysis centers on housing units where administrative records data indicated a vacant unit and the 2010 Census showed an occupied unit or vice versa.

Key Words: Administrative Records, nonresponse, census

Introduction

For the 2020 Census planning, Mule and Keller (2014) laid out the many issues and different potential ways that administrative records (AR) could be used in an adaptive way in the Nonresponse Followup (NRFU) operation. The Census Bureau implemented tests in 2013, 2014 and 2015 that used AR to reduce the number of contacts during the NRFU operation.

- Walejko et al. (2014) documented an adaptive design pilot test in October 2013 conducted in Philadelphia, Pennsylvania. The pilot test was of a small sample of addresses that were in the NRFU universe in the 2010 Census. This was the first step to test the feasibility of using AR to reduce the number of contact attempts during NRFU.

- The 2014 Census Test was conducted with a Census Day of July 1, 2014 in parts of Montgomery County, Maryland and the District of Columbia. Keller et al. (2016) documented how basic rules were developed to identify occupied and

[1] The views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

vacant addresses through the use of AR. One of their findings was that improvements could be made by using predictive modeling approaches as compared to rules.

- The Census Bureau conducted research and developed predictive modeling approaches that used logistic and multinomial regression predictions. Linear optimization approaches were then applied to maximize the AR determination given constraints. This new approach was implemented in the 2015 Census Test in Maricopa County, Arizona (Morris et al. 2016).

In addition to the mid-decade census tests discussed above, the development of possible AR models has been guided by comparing models retrospectively against 2010 Census results. For example, running a simulation on 2010 Census data, we counted how many addresses identified as vacant by the AR model were actually vacant during the 2010 census. Essentially, this type of analysis treats 2010 Census results as 'truth'. However, a difficulty underlying the evaluation of AR modeling is the inherent error in census results. Although the analysis using the 2010 Census results as 'truth' provides a solid basis for assessing model performance, it is not the only way model performance can be measured. It is possible that census quality could be improved using AR data that is not reflected by solely comparing the modeling results against 2010 Census 'truth'.

The goal of this paper is to provide greater detail about the differences between results from AR modeling simulations and the 2010 Census. Section 2 discusses how AR models have been developed and how AR data would be incorporated into the NRFU operation. Section 3 provides an example simulation with AR data and compares the results to what was observed in the 2010 Census Coverage Measurement program. Section 4 digs into some of the discrepancies seen between the AR and modeling results by looking more deeply at the AR data.

2. Administrative Records Modeling for NRFU

For the 2015 Census Test conducted in Maricopa County, Arizona and the 2016 Census Test conducted in Harris County, Texas and Los Angeles County, California, the Census Bureau identified occupied and vacant units using AR data and models. In this paper, we describe a national-level application of the same models that we applied during the 2016 Census Test. For the simulation in Section 3, we used the 2016 methodology to fit our AR models on a sample of the 2010 Census NRFU universe. We then applied the fit to the entire 2010 Census NRFU universe. See Morris et al. (2016) for specific details about the modeling approach and dependent and independent variables. Following the modeling, the NRFU address universe was split into four categories:
(1) units identified as occupied using AR (AR Occupied)
(2) units identified as vacant using AR (AR Vacant)
(3) addresses identified as having a non-existent housing unit (HU) using AR (AR Delete)
(4) addresses for which the AR data did not meet the quality thresholds for making an AR determination (No Determination).

2.1 Nonresponse Followup Contacts

This section gives an overview of the NRFU contact strategy related to enumerating some addresses with AR. This strategy was laid out in the release of the 2020 Operational Plan (U.S. Census Bureau 2015) and implemented in the 2016 Census Test. For the 2016 Census

Test, before the NRFU operation began, a NRFU address may have received up to four mailings before and after Census Day. These mailings included a letter encouraging the household to respond on the internet, two postcard reminders, and a paper questionnaire. If the address did not respond to these, a decision was then made about how many times to contact the address during the NRFU operation.

Figure 1 shows the flowchart of the contact strategy related to AR cases for the NRFU operation. Addresses determined to be AR Vacant received no contacts during the NRFU operation. While these units did not receive any NRFU visits, a postcard was mailed to them during the 2016 Census Test at the beginning of the NRFU operation. This allowed people at occupied addresses to self-respond by going online and filling out the internet questionnaire or dialing the questionnaire assistance phone number.

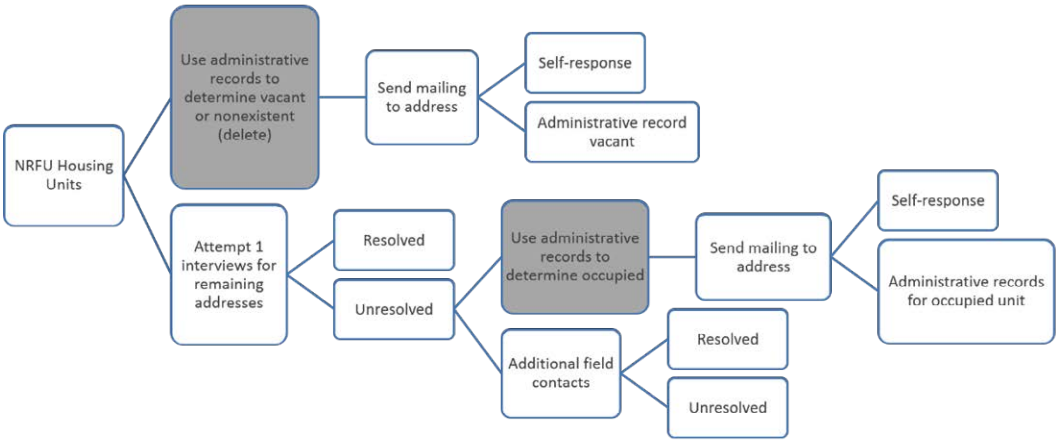

Figure 1: Nonresponse Followup Contact Strategy for Administrative Record Cases

The remaining cases received an initial field visit. This visit allowed each case to be resolved in several ways. It was resolved by

- completing the interview with the household member,
- determining the address to be vacant, or
- determining the address was not a HU.

If nobody in the household was home, the enumerator left a notice of visit. This notice of visit included information that instructed persons in the household to respond by going online, dialing the questionnaire assistance number, or sending the paper questionnaire that they received earlier.

Cases determined to be occupied by AR only received the initial visit in the 2016 Census Test. While they received only the initial visit, an additional postcard mailing was sent to the address. This postcard had information detailing how the household could still go online or dial the questionnaire assistance number to self-respond. As shown, there are several ways before and during NRFU that the Census Bureau is attempting to obtain and use self-responses before having to use AR determinations. The remainder of this paper focuses on coverage ramifications when applying the models to determine the AR Vacant, AR Delete, and AR Occupied cases that are shaded in gray in Figure 1.

3. Administrative Records Simulation

To identify vacant units with AR, we developed a multinomial logit model which predicted the probability that an AR address would have been enumerated as vacant during the 2010 Census. Independent variables in the model included variables indicating whether the census mailings could be delivered to the address and whether the AR sources indicated anyone lived at the address. The dependent variable had three possible values for each AR address in the NRFU universe:

- occupied
- vacant, or
- delete (i.e., not a HU).

We defined a Euclidian vacant distance function for AR Vacant identification as:

$$d_{AR_{Vac}} = \sqrt{(1 - p_{vacant})^2 + \left(0 - p_{occupied}\right)^2}$$

The formula shows that cases with the smallest distance were those with the highest vacant probability and lowest occupied probability. Starting with the smallest vacant distance, AR Vacant cases were identified by allowing for increased vacant distance values up to a threshold. This threshold was based on analysis of 2010 Census NRFU data. This modeling approach identified 5.048 million AR Vacant units nationally.

We defined a Euclidian delete distance function for AR Delete identification as:

$$d_{AR_{Del}} = \sqrt{(1 - p_{delete})^2 + \left(0 - p_{occupied}\right)^2}$$

The formula shows that cases with the smallest distance were those with the highest delete probability and lowest occupied probability. Starting with the smallest delete distance, AR Delete cases were identified by allowing for increased delete distance values up to a threshold. This threshold was based on analysis of 2010 Census NRFU data. This modeling approach identified 60 thousand AR Delete units nationally.

Two models were developed to identify AR Occupied units: a person-place model and a household (HH) composition model. Independent variables in the occupied models included variables indicating which AR sources placed people at the address and whether these people were found at a different address in the AR sources. The person-place model predicted the probability that an AR person would be enumerated at the sample address if fieldwork was conducted. The dependent variable was whether the AR person was at the address in the 2010 Census. The HH composition model predicted the probability that the sample address would have the same HH composition determined by NRFU fieldwork as its pre-identified AR HH composition. HH composition is defined by the number of adults in the unit and the absence or presence of children. The dependent variable was whether the 2010 Census HH composition.

Similar to AR Vacant and AR Delete, we defined a Euclidian occupied distance function for AR Occupied identification as:

$$d_{AR\_Occ} = \sqrt{\left(1 - p_{person-place}\right)^2 + \left(1 - p_{HH\ composition}\right)^2}$$

The formula shows that cases with the smallest occupied distance were those where the person-place probability was closest to one and the household composition probability was closest to one (i.e. the (1,1) point). Starting with the smallest occupied distance, AR Occupied cases were identified by allowing for increased occupied distance values up to an occupied threshold. This threshold was based on analysis of the 2010 Census NRFU data. This modeling approach identified 7.029 million AR Occupied units nationally.

Table 1 shows the distribution of cases identified as vacant and occupied by AR models and those for which no AR determination was made.

Table 1: NRFU Universe by AR Model Category

AR Model Category	Total	No Determination	AR Occupied	AR Vacant	AR Delete
N	49,817,252	37,632,033	7,077,460	5,047,583	60,176
Percent	100.0%	75.5%	14.2%	10.1%	0.1%

3.1 Comparing AR Modeling Simulation to 2010 NRFU

To understand the possible error in the model, we compared AR enumerations to the 2010 Census enumerations. That is, how many AR Occupied cases were occupied during 2010 Census? Similarly, how many AR Vacant cases were vacant during the 2010 Census? Table 2 shows three 2010 Census status outcomes:

- occupied,
- vacant, or
- delete.

Of the 5,107,759 cases identified as vacant or delete by AR, about 11.0% were classified as occupied in the 2010 NRFU operation. Similarly, of the 7,077,460 AR Occupied cases, 6.1% were classified as vacant and 1.4% were deleted.

Table 2: NRFU Status Assigned Via Simulation versus 2010 NRFU Status

AR Model Category	Total	2010 Census Status			%		
		Occupied	Vacant	Delete	Occupied	Vacant	Delete
AR Vacant	5,047,583	556,194	3,994,672	496,717	11.0%	79.1%	9.8%
AR Occupied	7,077,460	6,546,266	433,392	97,802	92.5%	6.1%	1.4%
AR Delete	60,176	6,301	15,874	38,001	10.5%	26.4%	63.1%

At the core of this paper is the idea that solely comparing possible AR modeling methods to previous 2010 Census results is insufficient because census results have errors. One might be tempted to conclude that the 556,194 units identified as vacant by AR but enumerated as occupied by NRFU in Table 2 are all misclassification errors attributed to the AR models. However, it is possible that all or some people in these units may be erroneous enumerations or whole-person census imputations. Hence, the AR simulation may be more accurately viewed through the prism of the 2010 Census Coverage Measurement (CCM) program. To understand this, Section 3.2 integrates potential AR modeling methods with the results from the 2010 CCM. Note that this paper focuses on one specific simulation as a qualitative demonstration. However, the use of CCM to evaluate AR models has been extended to many simulations.

3.2 Comparing AR Modeling Simulation to Census Coverage Measurement

The 2010 CCM program evaluated coverage of the 2010 Census to aid in improving future censuses. The CCM measured the net coverage and components of census coverage of HUs

and persons, excluding group quarters and persons residing in group quarters. The CCM sample design was a probability sample of 170,000 HUs. Remote areas of Alaska were out of scope for the CCM.

The general estimation approach for the components of census coverage categorized people into four categories:

- estimates of correct enumerations
- estimates of erroneous enumerations
- tabulations of whole-person census imputations
- estimates of omissions

Keller and Fox (2012) provided the 2010 components of census coverage for the national household population. Since a goal of the CCM process was to aid in improving future censuses, we show coverage properties of the AR simulation above to provide additional insight into the quality of the simulation.

Table 3 displays separate estimation domains for No Determination, AR Occupied, and AR Vacant or Delete cases. The AR Vacant or Delete cases are combined due to the small size of the AR Delete universe. It shows the components of census person coverage for the 7.077 million AR Occupied units, 5.048 million AR Vacant units, 60 thousand AR Delete units, and the remaining No Determination units that the AR models indicated as insufficient to enumerate with AR. The first column shows the census count. The census count is then broken into rates of correct enumeration, erroneous enumeration by duplication, erroneous enumeration for other reasons, and whole-person imputation.

Table 3 shows that the 2010 Census enumerated 16.243 million persons in the 7.077 million HUs the simulation identified as occupied by AR. Among these enumerations, 91.6% were estimated to be correct enumerations, 2.2% of these enumerations were erroneous due to duplication, 0.6% of these enumerations were erroneous due to some other reason, and 5.7% of these enumerations were whole-person census imputations. It is clear that not every census enumeration in these units was correct. Central to the point of this paper, a lower AR simulation total in comparison to the Census 2010 total may result in greater census quality given that 450 thousand persons were enumerated in error in the 2010 Census. In addition, 918 thousand persons had each characteristic imputed. In practice, calling these units occupied and enumerating them from AR data would result in no whole person imputations due to the presence of characteristic data such as age, sex, and date of birth for persons enumerated via AR.

In the 2010 Census, the 2010 Census enumerated 987 thousand persons in 5.108 million HUs that were classified as vacant or delete by AR models (see Table 3). However, not all these persons were correct enumerations. Among these enumerations, 70.7% were estimated to be correct enumerations, 8.5% of these enumerations were erroneous due to duplication, 1.5% of these enumerations were erroneous due to some other reason, and 19.3% of these enumerations were whole-person census imputations. In practice, if these units were determined to be vacant from AR data, AR methods would miss 698 thousand correctly enumerated persons.

Table 3: Components of Census Coverage by AR Simulation Category

AR Status	Census Count (Thousands)	Correct (%)	Erroneous (%)		Whole-Person Imputations (%)
			Duplication	Other	
U.S. Total	300,703	94.7	2.8	0.5	2.0
	(0)	(<0.1)	(<0.1)	(<0.1)	(0)
No Determination	283,473	94.9	2.9	0.5	1.7
	(0)	(0.1)	(0.1)	(<0.1)	(0)
AR Occupied	16,243	91.6	2.2	0.6	5.7
	(0)	(0.2)	(0.2)	(0.1)	(0)
AR Vacant/Delete	987	70.7	8.5	1.5	19.3
	(0)	(1.1)	(1.1)	(0.4)	(0)

Standard errors are shown in parentheses below the estimate. See Imel et al. (2013) on how CCM standard errors were estimated.
The 2010 Census count excludes persons in group quarters and persons in Remote Alaska.

4. Analysis Using Administrative Record Data

Section 3 looked at the results of the simulation by comparing to the 2010 CCM program. Section 4 analyzes the differences between the AR simulation and 2010 Census by delving into the AR simulation results and person-level AR data. To begin, the pre-imputed 2010 Census response file was processed and each of the 300,708,215 2010 Census household person records were eligible to be assigned a protected identification key (PIK) as part of the Person Identification Validation System (PVS). PIK assignment means that the person record was given a unique person number. Census person records were assigned a PIK in a cascading search through the four search modules discussed in Wagner and Layne (2014); geographic search, name search, date of birth search, and household composition search.

Each module has its own set of user defined blocking passes and parameter score thresholds. Records failing a module proceed to the next module. The quality of the match is then indicated by the search module used, blocking pass, and match score coming out of the PVS. Layne, Wagner, and Rothhaas (2014) examine the error in PIK assignment by the PVS associated with each of those search modules. In addition, all AR sourced data has a PIK assigned. For AR sources with a Social Security Number data field, each person record first goes through a verification module before the four search modules discussed above.

Sections 4.1 through 4.4 look at cases in which the AR simulation concluded the unit was vacant or delete, but the 2010 Census found it was occupied. Section 4.5 looks at the opposite construct - cases in which the AR simulation concluded the unit was occupied, but the 2010 Census found it was unoccupied.

4.1 Analysis of Vacant Distance for AR Vacant Cases

AR Vacant units that are in truth occupied introduce an undercount in the census. Undercount in the decennial census has been observed for young children and well as certain racial and ethnic minorities. The subsequent figure compares the vacant distance for all units identified as AR Vacant. Specifically, the box plot in Figure 2 shows the distribution of the vacant distances for the HUs classified as AR Vacant by 2010 Census

status. If the model is working as intended, the lower vacant distances for the AR Vacant cases would be associated with units that were vacant in the 2010 Census.

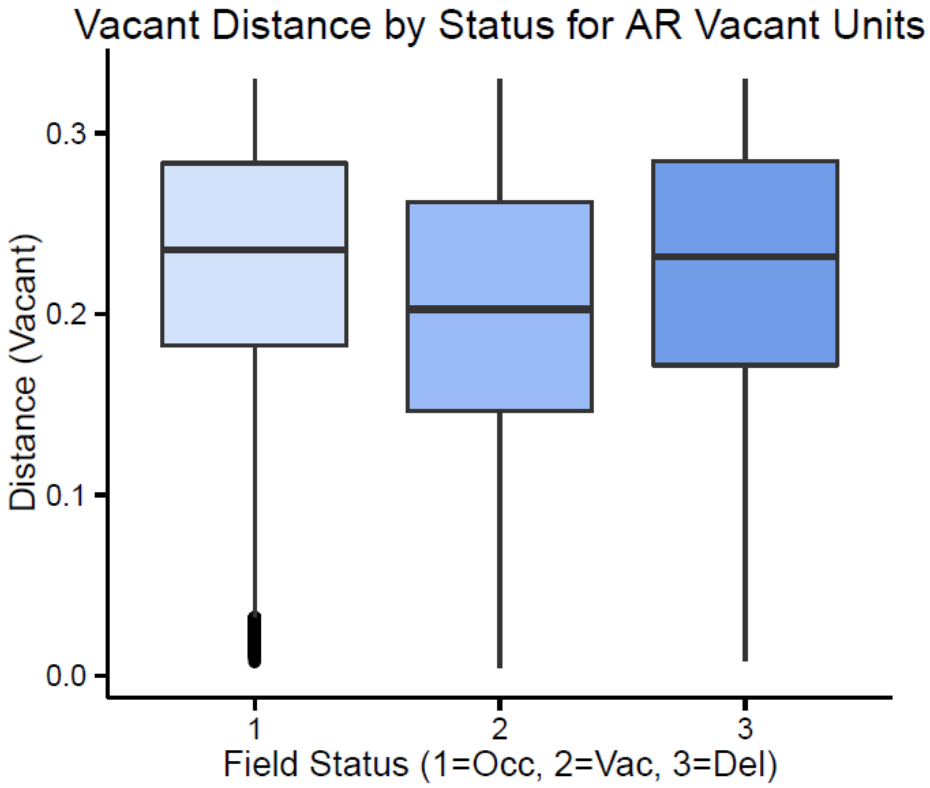

Figure 2: Vacant Distance by 2010 Census Status for AR Vacant Status

Figure 2 shows that the distance for AR Vacant units that were vacant in the 2010 Census is lower than that of the cases that were occupied or delete in the 2010 Census. Hence, the results suggest that if we were to lower the AR Vacant distance threshold, we would have proportionally more AR Vacant cases that were also vacant in the 2010 Census.

4.2 Analysis of Delete Distance for AR Delete Cases

AR Delete units that are in truth occupied also introduce an undercount in the census. In addition, the undercount also impacts the count of the HU inventory as an AR Delete unit would not be classified as a HU. The subsequent figure compares the delete distance for all units identified as AR Delete. Specifically, the box plot in Figure 3 shows the distribution of the delete distances for the HUs classified as AR Delete by 2010 Census status. If the model is working as intended, the lower delete distances for the AR Delete cases would be associated with units that were delete in the 2010 Census.

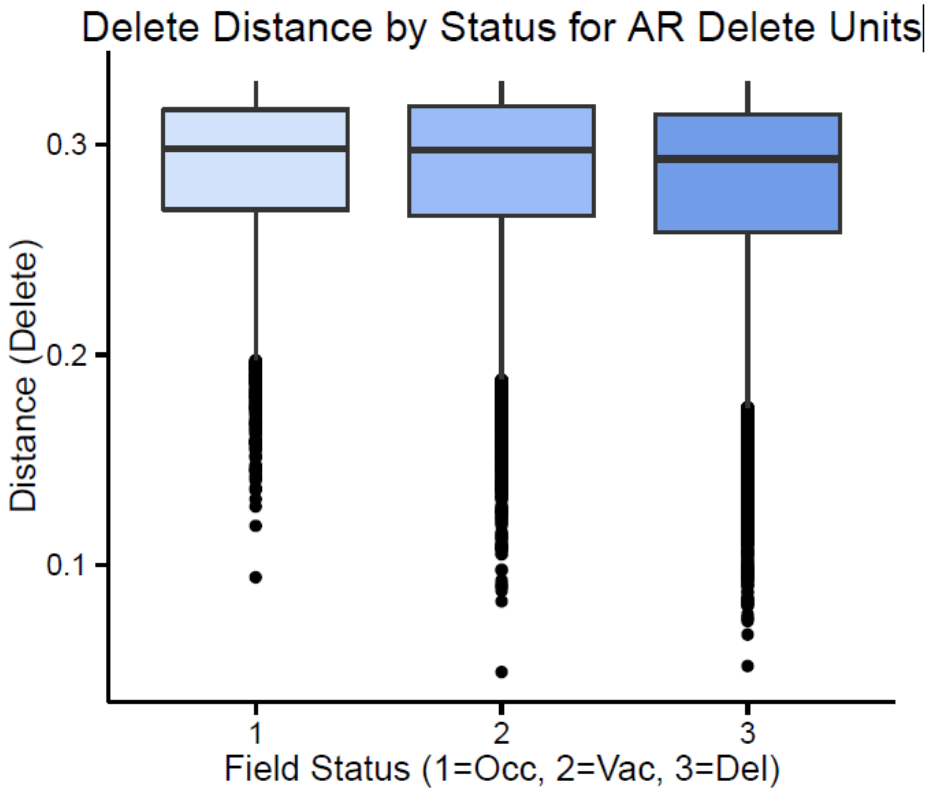

Figure 3: Delete Distance by 2010 Census Status for AR Delete Status

Figure 3 shows that the distance for AR Delete units that were delete in the 2010 Census is marginally lower than that of the cases that were occupied or vacant in the 2010 Census. Hence, the results somewhat indicate that if we were to lower the AR Delete distance threshold, we would have proportionally more AR Delete cases that were also delete in the 2010 Census.

4.3 AR Vacant or AR Delete and Census Occupied – What is the Completeness of Person Data?

Table 4 provides a comparison of AR Vacant and AR Delete HUs which were occupied in the 2010 Census against all occupied units in the 2010 Census. The table shows the availability of characteristic data for people in these units. The argument is that a lack of complete data for the units identified as AR Vacant and AR Delete but occupied in the census indicates that there is less certainty about the census enumeration. Complete data is reflected by the ability to assign a PIK as well as characteristics reported for the census person. Table 4 shows the rate at which all persons in the unit were able to be PIKed for all census persons, persons in units simulated to be AR Vacant, and persons in units simulated to be AR Delete. It also shows the rate at which characteristics were reported during the 2010 Census for the three groups. Note the difference between the AR Vacant and AR Delete columns in contrast to the 2010 Census.

Table 4: AR Vacant and AR Delete and 2010 Census Occupied vs. 2010 Census Persons

	AR Vacant	AR Delete	2010 Census
Total HUs	556,194	6,301	116,716,292
Total Persons	975,136	12,127	300,758,215
% Persons PIKed	48.8	16.4	90.7
% Persons with Age reported	52.2	27.5	89.6
% Persons with Sex reported	79.3	65.5	96.5
% Persons with Race reported	74.3	60.4	94.0
% Persons with Hispanic reported	73.7	60.4	93.3
% Persons with Relationship reported	77.5	62.9	96.0
% Persons with No Information	19.1	33.0	1.9
% All Persons in HU PIKed	47.6	19.2	87.1

For the 556,194 AR Vacant units that were occupied in the 2010 Census, there were 975,136 census persons. We were able to assign a PIK to 476,078 (48.8%) of these cases. For the 6,301 AR Delete units that were occupied in the 2010 Census, there were 12,127 census persons. We were able to assign a PIK to 1,989 (16.4%) of these cases. By comparison, of the 300,758,215 person enumerations in the 2010 Census, 272,911,055 were able to be PIKed (90.7%). With respect to age, 52.2% of census persons in AR Vacant units reported an age while 27.5% of census persons in AR Delete units reported an age. This is compared to 89.6% of all 2010 Census persons reporting an age. Going down the rows, the general picture is that characteristic reporting is lower by about 20-30 percentage points for the AR Vacant or AR Delete persons when compared to the overall 2010 person rate. For the AR Vacant or AR Delete HUs, we observe higher rates of persons with no characteristic information and lower rates where everybody in the HU is assigned a PIK.

4.4 AR Vacant or AR Delete and Census Occupied – Was there evidence of these people in another location?

Using this PIK information, we look into our AR data and determine where we had evidence of these persons, if at all. The idea is that, if these census persons could be found somewhere else in AR, it could be that the occupied designation in the Census was introducing duplication. In Table 5, we see that 293,945 of the 476,078 (61%) PIKed AR Vacant persons were found in AR at another address, and 424 of the 1,989 (21%) PIKed AR Delete persons were found in another address in AR.

Table 5: AR Vacant and AR Delete, 2010 Census Occupied Persons

2010 Census Occupied	AR Vacant	AR Delete
Total Persons	975,136	6,301
Total PIKed Persons	476,081	1,986
Total PIKed Persons Elsewhere in AR	293,945	424

4.5 AR Occupied and Census Unoccupied

AR Occupied units that are in truth unoccupied introduce an overcount in the census. For this section, we analyze whether these AR persons are in units that were indeed occupied in the 2010 Census. Table 2 shows that there were 433,392 AR Occupied units classified as vacant and 97,802 AR Occupied units classified as delete in the 2010 Census. We analyze the properties of these units in administrative records that lead us to believe that these are occupied units. To do so, we use the 6,546,266 AR Occupied units classified as occupied for comparison.

The subsequent figure compares the occupied distance for all units identified as AR Occupied. Specifically, the box plot in Figure 4 shows the distribution of the occupied distances for the HUs classified as AR Occupied by 2010 Census status. If the model is working as intended, the lower occupied distances for the AR Occupied cases would be associated with units that were occupied in the 2010 Census.

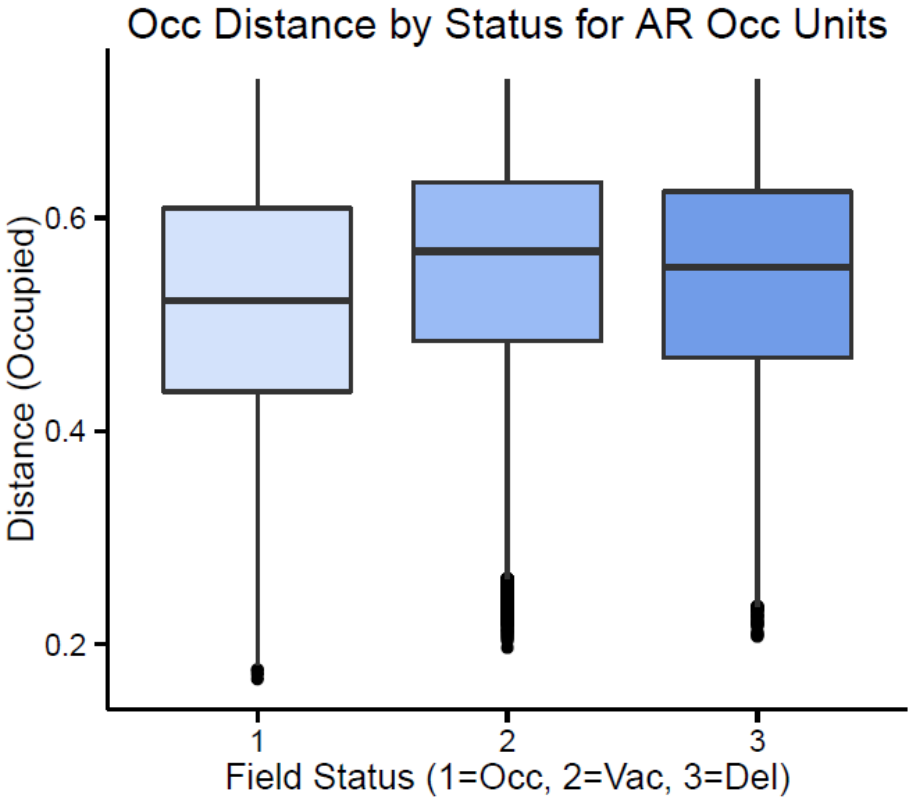

Figure 4: Occupied Distance by 2010 Census Status for AR Occupied Status

Figure 4 shows that the distance for the 6.5 million AR Occupied units that were occupied in the 2010 Census is lower than that of the cases that were vacant or delete in the 2010 Census. Hence, the results suggests that if we were to lower the AR Occupied cutoff, we

would have proportionally more AR Occupied cases that were also occupied in the 2010 Census.

We look further at these AR persons that were in unoccupied HUs in the 2010 Census. Table 6 shows that there were 433,392 AR Occupied units classified as vacant. In these units, 818,982 AR persons were identified. Similarly, there were 97,802 AR Occupied units classified as delete in the 2010 Census. In these units, 204,072 AR persons were identified

Table 6 shows the age distribution of the cases simulated as AR Occupied and found to be vacant or delete in the 2010 Census. These distributions are compared to the overall distribution of all 2010 Census persons.

Table 6: All 2010 Census Occupied Units vs. AR Occupied/Field Unoccupied Units

2010 Census Occupied	2010 Census	AR Persons in Vacant Units	AR Persons in Delete Units
Total HU	116,716,292	433,392	97,802
Total Persons	300,758,215	818,982	204,072
% Age 0-4	6.7%	8.6%	8.0%
% Age 5-9	6.8%	7.7%	8.6%
% Age 10-17	11.1%	9.2%	10.5%
% Age 18-29	15.9%	15.9%	12.6%
% Age 30-49	27.3%	27.1%	30.6%
% Age 50+	32.2%	31.6%	29.8%

Keller and Fox (2012) show a significant undercount of age 0 to 4 persons in the 2010 Census. Hence, the fact that the age distribution for AR persons in vacant and delete units has a larger proportion of 0 to 4 year olds than the 2010 Census distribution might indicate that the AR simulation has merit. Similarly, Keller and Fox (2012) show a significant overcount of age 10 to 17 persons, 50+ females, and 50+ males in the 2010 Census. Consequently, the fact that the age distribution for AR persons in vacant and delete units has a smaller proportion of 10 to 17 year olds, 50+ females, and 50+ males than the 2010 Census distribution might indicate that the AR simulation has value.

5. Discussion

For the 2020 Census, research involving the integration of AR in the NRFU operation is ongoing. The goal is to save costs and maintain data quality. Retrospective simulations that incorporate AR data into the 2010 Census have been critical to understand the effect on populations and missing data. While comparing back to the 2010 Census, it is necessary to understand that the 2010 Census has errors. Hence, not all differences between AR simulations and 2010 Census results are unquestionably errors in the AR modeling process. This paper further investigates the sources of those differences and sheds light on why discrepancies occur between AR and the 2010 Census. This process informs the further refinement of the AR modeling process.

6. References

Imel, L., Mule, V.T., Seiss, M., and Mulligan, J. (2013), "2010 Census Coverage Measurement Estimation Methods: Measures of Variation," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-J-06.

Keller, A. and Fox, T. (2012), "2010 Census Coverage Measurement Estimation Report: Components of Census Coverage Results for the Household Population in the United States," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-04.

Keller, A., Fox, T., and Mule, V.T. (2016). "Analysis of Administrative Record Usage for Nonresponse Followup in the 2014 Census Test." U.S. Census Bureau.

Layne, M., Wagner, D., and Rothhaas, C. (2014). "Estimating Record Linkage False Match Rate for the Person Identification Validation System." Center for Administrative Records Research and Applications Working Paper #2014-02. Washington, DC: U.S. Census Bureau.

Morris, D.S., Keller, A., and Clark, B. (2016). "An approach for using administrative records to reduce contacts in the 2020 Decennial Census," Statistical Journal of the International Association of Official Statistics, 32 (2016): 177-188.

Mule, V.T. and Keller, A. (2014), "Using Administrative Records to Reduce Nonresponse Followup Operations," in JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3601-3608.

U.S. Census Bureau 2015. *2020 Census Operational Plan*. Washington DC: Census Bureau. Available at: http://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan.pdf. (accessed September 2017.

Wagner, D. and Layne, M. (2014). "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software." Center for Administrative Records Research and Applications Working Paper #2014-01. Washington, DC: U.S. Census Bureau.

Walejko, G., Keller, A., Dusch, G., and Miller, P.V. (2014). "2020 Research and Testing: 2013 Census Test Assessment," U.S. Census Bureau.