Administrative Records Planning and Research for the National Advisory Committee Fall 2017 Meeting, November 2, 2017

1. Introduction

The goal of the 2020 Census is to count each person only once and in the correct location. This is to be achieved at a lower cost per household (adjusted for inflation) than the 2010 Census while maintaining the same data quality. To meet this goal, the Census Bureau has researched fundamental changes to the design, implementation, and management of the 2020 Census. The Census Bureau laid out its methodology for achieving this task in the 2020 Operational Plan (U.S. Census Bureau 2016). One of the major innovation research areas for the 2020 Census has been developing approaches to incorporate administrative records (AR) and third-party data into the census design. AR are data maintained by external government entities, and third-party data are sources compiled by private organizations. These data are obtained by the Census Bureau to aid in the census and surveys from organizations such as the Internal Revenue Service (IRS), Social Security Administration (SSA), and the United States Postal Service (USPS).

This document provides an overview of our methodology for using AR in the Nonresponse Followup (NRFU) operation for the decennial census. In the 2010 Census, the NRFU operation encompassed census enumerators visiting areas to verify the status for every address that received a 2010 Census questionnaire but did not respond by mail. NRFU included about fifty million addresses requiring up to six enumerator visits each, totaling about $1.6 billion (Walker et al. 2012). Following NRFU, each of the previously unresolved 50 million addresses was assigned one of three possible housing unit statuses: occupied, vacant, or nonexistent. If the unit was determined to be occupied, a housing unit count was determined and each person was rostered with basic demographic characteristics such as name, age, date of birth, race, Hispanic origin, and relationship to the householder.

We have developed a methodology to use administrative records to determine the status of NRFU addresses (i.e., occupied, vacant, or nonexistent) and, when determined to be occupied, to enumerate the address with the occupants and many of their characteristics. This methodology uses administrative records only where they are strongest, thus allowing a full field contact strategy to be conducted for the remaining NRFU addresses. The remainder of this document describes the method and provides some quality metrics used in our research using 2010 Census data. We also provide more information about how characteristics will be obtained for the administrative record enumerations. Based on this work, the characteristic imputation processing is considering the usage of these sources if a respondent does not provide the information in their response.

2. Administrative Records Modeling Methodology

The overarching principle governing AR use during the NRFU operation has been that the Census Bureau will selectively use AR to determine the status and population count of a subset of addresses in the NRFU universe. The development of the AR models we will potentially be using in the 2020 Census has been guided by two processes. First, we have been testing our approaches in census tests conducted from 2013 to 2016 and culminating with our 2018 End-to-

End Census Test. Second, we have performed an internal validation by comparing the outputs of our approach to 2010 Census results.

The list of addresses for the 2020 Census is taken from the Master Address File (MAF). Each address receives mailings in accordance with the Census Bureau's mailing strategy. Most addresses will receive a letter with an invitation to complete the census on the Internet, followed by a reminder letter a few days later. In some areas, a paper questionnaire will be included in the first mailing. If an address has not responded after these first two mailings, additional reminder postcards along with a paper questionnaire will be sent to the address. If the address still does not respond to these mailings, then a decision needs to be made about how that address will be handled in the NRFU operation.

Each address in the NRFU universe will be evaluated under specific, pre-identified criteria to determine if and how AR can be applied to the address in the NRFU operation. We begin by building rosters for each address using persons from the following data sets: Internal Revenue Service (IRS) 1040 Individual Tax Returns, IRS 1099 Information Returns, the Center for Medicare and Medicaid Services (CMS) Medicare Enrollment Database, and Indian Health Service (IHS) Patient Database. The exclusive use of federal sources to create AR rosters is motivated by the concerns about the use of third party data expressed in Steinhardt et al. (2014).

We supplement address-level data with information from the United States Postal Service (USPS). We also include other information maintained by the Census Bureau about the address. We supplement person-level data with information from other AR sources such as the Social Security Administration (SSA). Last, we enrich the address- and person-level data with block-level information using data from prior censuses and surveys like the American Community Survey.

Our approach utilizes information from the USPS about whether our mailings around Census Day were successful. We also incorporate whether persons taken from our AR roster are found at the addresses. We apply criteria to each address to determine whether the unit was vacant or did not meet the Census Bureau's definition of a housing unit (i.e. nonexistent) on Census Day.

From this we can make a determination about the initial status of the address. For example, suppose a mailing was returned by the USPS and coded as "vacant." In addition, suppose no people were found at the address from the AR sources. Both pieces of information increase the likelihood that we will make the determination that the address is vacant without doing any field work. Similarly, suppose a mailing was returned by the USPS and coded as "No Such Number" and no people were found at the address in the other AR sources. These pieces of information increase the likelihood that we will make the determination that the address is nonexistent without doing any field work. In essence, in order for a unit to be initially identified as vacant or nonexistent, the business rules require that a mailing around Census Day has been returned to the Census Bureau with a reason why the USPS could not deliver it.

Our second application of AR is for occupied households that have not responded to our multiple mailings by either answering the census via the Internet or returning a census questionnaire. We apply criteria to each address to determine the eligibility of the address for enumeration by AR.

One general tenet is that the AR persons can be found only at the address of interest, i.e., they are not duplicated at other addresses. Another basic framework is stability of the AR roster across multiple years. These constructs increase the likelihood that we will make the determination that the address is occupied. If we are not confident that the AR data we have compiled for the address are accurate, then we will not reduce field contacts. Instead will send enumerators to conduct the full number of in-person visits just as we have done in the past.

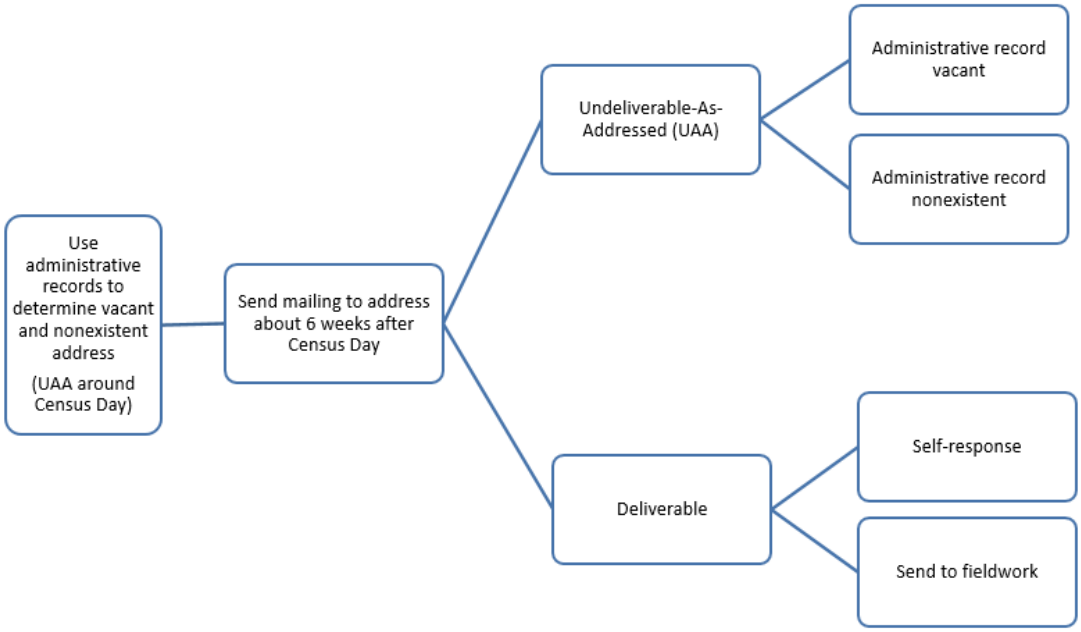
For each address to be eligible to be enumerated by AR, we also apply rigorous business rules to ensure that we have accurately captured data on the people actually living there. First, multiple sources must consistently identify the people living at the address. Their ages must be available, and we only attempt to enumerate households with six or fewer people living at that address. In addition, we only attempt to enumerate households with between one and three adults, where an adult is defined as any person age eighteen or older. We will only use the enumeration from administrative records if the household does not self-provide the information. Section 3 provides details about how we are providing these addresses opportunities to provide their census information.

3. Nonresponse Followup Contact Strategy with Administrative Records

The previous section discussed how AR are used to make determinations about the status of an address in the NRFU workload. This section discusses how these AR determinations are incorporated into the NRFU contact strategy for the 2018 End-to-End Census Test. The contact strategy has been updated throughout the decade based on findings from earlier census tests. Before NRFU begins, an address will be eligible to receive up to five mailings in March and April 2018. In general, the first mailing will encourage respondents to respond on the internet. Addresses that have not responded by a given date will receive a paper questionnaire that can be filled out and returned in the mail. If an address has not responded after all mailing attempts, the address will be included in the NRFU workload. Note that some areas with anticipated low internet response will receive a paper questionnaire earlier on during the self-response period.

For all addresses in the NRFU workload, a decision will be made about how many times to visit the address during the NRFU operation based on the quality of the AR data. Addresses initially determined to be vacant or nonexistent by administrative records are treated as shown in Figure 1. These addresses will have had one of the mailings near Census Day marked as undeliverable-as-addressed (UAA). When NRFU begins in mid-May, these addresses are not visited by an enumerator. Instead, we will send an additional mailing to these addresses. Only those cases for which this additional mailing is returned by the USPS as UAA will be determined to be vacant or non-existent based on AR. Thus, the AR vacant and nonexistent addresses must have had at least two mailings marked as UAA. One of them needs to be near Census Day and the other one is in mid-May. If the postal carrier is able to deliver the May mailing, then we will add the address back into the NRFU field workload. The address will then receive a full set of visits from an enumerator or be completed by self-response.

Figure 1. Contact Strategy for AR Vacant and Non-Existent Addresses

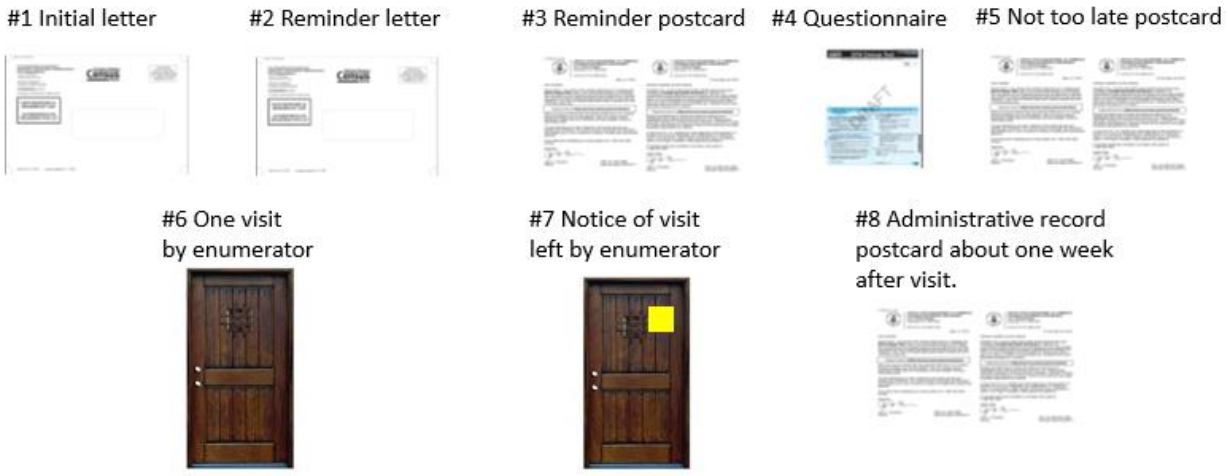


Addresses that are determined to be AR occupied are treated as shown in Figure 2. These addresses will have received several mailings prior to the start or NRFU. The first row of the figure shows the five mailings that an address will receive between the middle of March and the end of April. During NRFU, these addresses will receive a single visit by an enumerator. This allows the case to be resolved in several ways. Units could be resolved by a completed interview with a household member or by the NRFU enumerator determining that the address was vacant or nonexistent. Otherwise, the NRFU enumerator will leave a notice of visit at the address. The notice of visit includes information that allows the household to respond by going online, calling the Census Questionnaire Assistance (CQA) telephone number, or returning the paper questionnaire that was sent earlier.

For address not resolved by the enumerator on this first visit, we will then look to administrative records. Addresses with high-quality AR information will be sent an additional mailing to encourage self-response by providing information about going online or calling the CQA telephone number. If no self-response is received, then the AR occupied determination will be used to resolve the address. Addresses without high-quality administrative records will receive additional field contacts to attempt to resolve the case.

As shown, there are several ways before and during the NRFU operation that the Census Bureau will attempt to obtain and use self-responses before having to resolve cases using AR information. Further, we are only using AR information to resolve cases for addresses with high-quality AR information. Addresses without AR information and address with AR information that does not meet the quality thresholds will receive a full set of visits from a NRFU enumerator.

Figure 2. Contact Strategy for AR Occupied and No Determination Addresses



Note: Some addresses may receive the questionnaire in the first mailing.

4. Results from the 2016 Census Test

The 2016 Census Test was conducted in portions of Los Angeles County, California and Harris County, Texas. This section summarizes the AR modeling results for the 2016 Census Test. In order to evaluate the results of the AR determinations, we selected an evaluation sample of AR determinations to remain in the NRFU universe, where they were sent to the field. The subset was selected by sorting by how confident we were in the determination from most to least and systematically selecting every fifth unit. The resolution of these units by NRFU fieldwork was used to assess the accuracy of our AR determinations. That being said, many of the cases in the evaluation sample were not resolved by the NRFU fieldwork. This limited our ability to assess the AR determinations.

Table 1 compares the AR determinations to the information collected in NRFU fieldwork for the evaluation sample. Overall, the AR occupied, Phase 1 units had about 80 percent status agreement, AR vacant had about 43 percent status agreement, and AR nonexistent units had about 49 percent status agreement. Note that each of these classifications was also associated with an unresolved rate near or above 10 percent. For the AR vacant and AR nonexistent cases, the percent of cases that were resolved as occupied was higher than expected. This result led us to update the contact strategy for cases with AR vacant and nonexistent determinations, as described in Section 3.

Table 1. Housing Unit Status for AR Determinations in the Evaluation Sample

	Total	Occupied		Vacant		Nonexistent		Unresolved	
AR Determination	N	%	SE	%	SE	%	SE	%	SE
AR Occupied, Phase 1	2,338	80.3	0.7	4.8	0.4	1.7	0.2	13.2	0.6
AR Vacant	715	21.1	1.3	42.8	1.6	20.7	1.2	15.4	1.2
AR Nonexistent	313	29.1	2.1	10.9	1.4	48.6	2.2	11.5	1.7
AR Occupied, Phase 2	64	51.6	5.9	4.7	2.2	1.6	1.4	42.2	5.7

5. Quality Evaluation Using 2010 Data

As part of our research, we have attempted to quantify the impact on quality based on the use of administrative records and third-party data to reduce contacts. The challenge is that there are multiple dimensions of quality to be considered. In this section, we show some examples of quality evaluations conducted. These metrics have fallen into two main types. The first is micro-level and summarizes the amount of agreement and differences for cases when comparing on an individual basis. The second is macro-level and shows the implications of our change on aggregate-level results.

In this section, we use a retrospective analysis of applying these approaches on the 2010 Census. For this analysis, we determined about 10 percent of the NRFU addresses to be AR vacant, about 15 percent to be AR occupied, and 0.1 percent to be AR nonexistent.

5.1. Micro-Level Evaluations of Quality

For the micro-level evaluations, we compare the results from the AR modeling to the 2010 Census results at individual addresses. Note that this retrospective analysis does not account for the aspects of the NRFU contact strategy discussed in Section 3. In particular, this analysis does not account for the single contact attempt for AR occupied cases nor the additional mailings for all AR vacant, nonexistent, and occupied cases. We treat all of the AR determination cases as if that is the final determination for the address.

For the about 10 percent of the NRFU eligible cases determined to be administrative record vacant, Table 2 shows the comparison of occupancy status of the administrative record cases to the 2010 Census NRFU outcomes. The table shows that 80 percent of the cases were both vacant in our determination and the Census NRFU outcome. There is 10 percent of the cases for which the Census result was occupied. One thought was that this could be a potential source of undercounts. In assessing this result, Cresce (2012) shows that the gross vacancy rates from the Housing Unit Vacancy Survey were higher than the 1990, 2000, and 2010 Census results. This indicates that the census has had a trend of understating the number of vacant addresses. Also, to potentially minimize undercoverage, the NRFU contact strategy shown in Section 3 includes an additional mailing during the Nonresponse Followup operation. This provides another opportunity for people that should be enumerated at these addresses to respond without being visited by an enumerator. Furthermore, only those addresses which are marked as UAA for this additional mailing will be assigned the AR vacant status. We anticipate that this additional UAA requirement will reduce the number of AR vacant cases that are actually occupied.

Table 2. 2010 Census NRFU Status Outcomes for Administrative Record Vacant Cases

2010 Census NRFU Outcome	Percent
Occupied	10.4
Vacant	79.8
Nonexistent	9.8

Table 3 shows the comparison of the administrative records nonexistent cases to the 2010 NRFU outcomes. This simulation identified only 0.1 percent of the NRFU addresses as AR nonexistent. About 61 percent of the AR nonexistent cases were actually nonexistent in the 2010 Census, and

another 26 percent were vacant in the census. For the 10 percent of AR nonexistent cases that were occupied in the 2010 Census, we anticipate that the additional UAA requirement for the NRFU mailing will reduce this rate. We will assess how changes to the address canvassing operation for the 2020 Census may impact these results.

Table 3. 2010 Census NRFU Status Outcomes for Administrative Record Nonexistent Cases

2010 Census NRFU Outcome	Percent
Occupied	10.3
Vacant	26.0
Nonexistent	61.4

For the about 15 percent of the NRFU eligible cases determined to be administrative records occupied, Table 4 shows the comparison of the occupancy status of the cases to the 2010 Census NRFU outcomes. The table shows that 89.9 percent of the time, the administrative records and NRFU interview determined that the address was occupied. Another 10 percent of the addresses were either vacant or determined to not exist. One part of the design that accounts for this finding is that the NRFU contact strategy does conduct one field visit for these cases. This field visit does allow the enumerator to determine that these cases are vacant or are nonexistent.

Table 4. 2010 Census NRFU Status Outcomes for Administrative Record Occupied Cases

2010 Census NRFU Outcome	Percent
Occupied	89.9
Vacant	8.4
Nonexistent	1.6

For the administrative record occupied cases, we also compared the population count based on the administrative record roster to the Census result. Table 5 shows that the population count agreed 62.1 percent of the time. For 22.3 percent of the cases, the administrative record count was higher than the census result. For 15.7 percent, the administrative record count was lower than the census count. Census results that were vacant or nonexistent had a population count of zero for these comparisons. These results show that when the counts differ that the administrative record counts were higher more often.

Table 5. Count Agreement for Administrative Record Occupied Cases

Count agreement	Percent
AR count higher	22.3
Same count	62.1
AR count lower	15.7

5.2. Macro-Level Evaluations of Quality

The micro-level summations of quality tell only one part of the story of how administrative records and third-party data would impact census quality. In assessing the macro-level implications, we did a retrospective analysis of the 2010 Census. In this simulation on the 2010 Census, we used administrative records and third-party data to enumerate about 5 million vacant

and about 7.5 million occupied housing units. Like the micro-level analysis, this simulation does not account for the aspects of the NRFU contact strategy discussed in Section 3.

One method of assessing the quality is to examine the extent of administrative records usage by different geographic areas. Here, we consider areas defined by concentrations of Hispanic and Non-Hispanic Black populations based on the 5-year ACS block group estimates. Table 6 shows the number of 2010 NRFU addresses by concentrations of the Hispanic population in the block group. The table also shows the percent of NRFU cases that were assigned the AR occupied and vacant determinations by concentrations of the Hispanic population. We see that areas with higher Hispanic concentrations have lower rates of AR occupied and vacant determinations. Thus, these areas have a higher rate of cases that receive the full contact strategy.

Table 6. 2010 Administrative Records Usage by Hispanic Population Concentration

ACS 5-Year Estimate of Percent of Block Group that is Hispanic	2010 NRFU Addresses (millions)	AR Determination (row percent)		
		AR Vacant (%)	AR Occupied (%)	Full Contacts (%)
0 to 10 percent	31.3	11.6	15.9	72.5
10 to 20 percent	6.8	9.4	15.7	74.9
20 to 30 percent	3.6	8.5	14.4	77.1
30 to 40 percent	2.2	7.6	13.3	79.1
40 to 50 percent	1.6	7.5	12.0	80.5
50+ percent	4.2	4.3	9.6	86.1
Total	49.8	10.1	15.0	74.9

Table 7 repeats this analysis for the Non-Hispanic Black population. As with the Hispanic concentrations, we generally see lower rates of AR usage in areas with higher Non-Hispanic Black concentrations. However, we do see a higher rate of AR vacant assignment in the areas where the Non-Hispanic Black population is estimated to be 50 percent or more of the population total. This result along with results from the mid-decade tests have led us to update the NRFU contact strategy for AR vacant and nonexistent cases as discussed in Section 3. For the AR vacant and nonexistent cases, we will require that the additional mailing in May be returned as UAA by the postal carrier

Table 7. 2010 Administrative Records Usage by Non-Hispanic Black Population Concentration

ACS 5-Year Estimate of Percent of Block Group that is Non-Hispanic Black	2010 NRFU Addresses (millions)	AR Determination (row percent)		
		AR Vacant (%)	AR Occupied (%)	Full Contacts (%)
0 to 10 percent	33.6	10.5	16.0	73.5
10 to 20 percent	5.4	8.2	15.6	76.2
20 to 30 percent	2.9	8.1	14.1	77.8
30 to 40 percent	1.8	8.3	13.0	78.7
40 to 50 percent	1.2	8.9	11.9	79.2
50+ percent	4.9	12.1	9.6	78.3
Total	49.8	10.1	15.0	74.9

6. Characteristic Assignment

If a self-response is not received then we will use the roster developed from the administrative record sources described earlier in Section 2 to enumerate the household. Table 8 shows the characteristics which need to be assigned at the person and housing unit level.

Table 8. Characteristics for the 2020 Decennial Census

Person-Level Characteristics	Housing Unit Characteristics
Age Date of Birth Sex Hispanic Origin Race Relationship to Householder	Tenure (given that the unit is occupied), Detailed vacancy (given that the unit is vacant)

Because we are not getting characteristics from an interview, we obtain the demographic characteristics of the household from various sources in a hierarchical manner. Where appropriate, we first use data collected by the Census Bureau in the prior census or other surveys. Next, where available, data from other administrative records sources are used. Last, third-party sources are used to determine the tenure status (owner or renter) of the household. Table 9 shows the administrative record and third-party data sources that have been researched to assign characteristics.

For age, date of birth, and sex, the Census Numident has been researched as second source behind other data obtained by the Census Bureau. The SSA Numident file contains all transactions ever recorded for any single SSN. The Census Bureau then builds a Census Numident on a regular basis from personal information derived from the SSA Numident file. All transactions related to a given SSN are resolved to produce a Census Numident file containing one data record for each SSN. See Luque and Wagner (2015) for more information on the Census Numident. The high quality of Census Numident age and sex data provides confidence in its use for characteristic assignment. With respect to age, the 2010 Census Match Study (Rastogi and O'Hara 2012) showed that previous census responses had a 95.7 percent match rate to current census responses. It also showed that Social Security Numident responses had a 97.9 percent match rate to current census responses. With respect to sex, the 2010 Census Match Study (Rastogi and O'Hara 2012) showed that previous census responses had a 99.5 percent match rate to current census responses. It also showed that Social Security Numident responses had a 99.4 percent match rate to current census responses.

For race and Hispanic origin, multiple AR sources have been analyzed as a second option to previous census data. First, the Census Numident has been researched. The file contains race and Hispanic origin data as well information on place of birth. In addition, we have researched using the Census Best Race and Hispanic Origin File. This file is developed by the Census Bureau's Center for Administrative Record Research and Application (CARRA). CARRA obtains information from multiple administrative record and third-party data sources on race and Hispanic origin. When the information is different, CARRA creates a series of business rules to identify a best race and Hispanic origin. More information can be found in Ennis et al. (2015). Last, we have looked at using state program participation data including Supplemental Nutrition

Assistance Program (SNAP), Women, Infants, and Children (WIC), and Temporary Assistance for Needy Families (TANF) datasets to aid in the assignment of race and Hispanic origin. This came out of a recommendation in Dowling et al. (2016) to use state records to obtain demographic administrative data.

For relationship to householder, we have researched using the CARRA Kidlink file. The CARRA Kidlink file links a child to the mother and father using the names of the parents from the Social Security number application for the child. More discussion of CARRA Kidlink is in Section 7. The purpose of using CARRA Kidlink was to assign a child relationship status given that parent was the householder. In addition to CARRA Kidlink, we have looked at using state program participation data (SNAP, WIC, and TANF) to aid in the assignment of householder and relationship to householder. Again, this came out of a recommendation in Dowling et al. (2016).

To assign a tenure status of owned or rented, we have used program participation data from Housing and Urban Development. This includes data from Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS). In addition, we have researched the use of commercial tax and deed information to help in this assignment of tenure to units which we identify as occupied through administrative records.

Table 9. Administrative Record and Third-Party Data Researched to Assign Characteristics

Characteristics	Data Sources
Age, Date of Birth, Sex	Census Numident
Race, Hispanic Origin	Census Numident Census Best Race and Hispanic Origin File State Program data (SNAP, TANF, WIC)
Relationship to Householder	Census KIDLINK State Program data (SNAP, TANF, WIC)
Tenure	Housing and Urban Development TRACS and PICs files Commercial Tax and Deed information

We have described using characteristic data collected by the Census Bureau from previous censuses and survey for cases enumerated by AR in 2020. After the legacy census data has been exhausted, then the federal or state AR data is used to assign characteristics. Last, tax and deed information collected by a third-party vendor is used for tenure assignment. Simultaneously, we also have been researching the use of previous census, survey, and AR data to aid with characteristic imputation for person records in need of item imputation but enumerated by self-response or NRFU. This helps us to avoid using techniques like hot deck to perform characteristic imputation. Instead, we can use the characteristics from previous census, survey, or AR data if that the 2020 Census person is linkable to administrative records,

7. Additional Administrative Records and Third-Party Sources

The Census Bureau continues to acquire and research additional administrative records and third-party sources that may improve our methodology. This section provides a brief discussion of some of these additional sources.

7.1. Center for Administrative Records Research and Applications Kidlink File

The CARRA Kidlink file is a research file create by the Census Bureau. The CARRA Kidlink research file is a child-to-parent linking dataset that is created using data from Social Security Number applications. For each child on the Social Security Administration Numerical Identification file, the CARRA attempted to assign a mother Protected Identification Key (PIK) and father PIK using the names of the parents from the SSN applications. A PIK is an encryption of a Social Security Number or Tax Identification Number (TIN) and is used as a unique identifier for matching person records between various AR and census sources. Note that for the child SSN applications used to build the CARRA Kidlink file, the SSNs were not available for the parent records. Then CARRA attempted to assign a PIK for the parents by matching the parent names to their set of reference files. Table 10 shows the success rate of assigning a PIK to the mother and the father of the child. The CARRA Kidlink filed used here includes births between 1996 and 2015. For about two-thirds of the children, a PIK could be assigned to both the mother and the father. Another 21 percent of children had only the mother PIK assigned. It may be that there was no father information available for these records.

Table 10. Parent PIK Assignment Rate

PIK assigned to…	Count	Percent
Total Children	85,133,839	100.0
Both Parents	56,471,835	66.3
Mother Only	17,757,808	20.9
Father Only	3,522,364	4.1
Neither Parent	7,381,832	8.7

The CARRA Kidlink file may help address a potential undercoverage of children in the core AR sources. While children can appear on any of the core sources, the vast majority of children are found only in the IRS 1040 source. IRS 1099 information returns are person-level records such as bank account interest statements and W-2 wage statements. These records typically apply to adults. Medicare generally covers the population above age 65. The CARRA Kidlink file may allow us to link additional children to an address via the parents that are found on the core AR sources.

We are researching possible uses of the CARRA Kidlink file in the AR modeling process, including:

- Adding the file to the set of core sources
- Using the data to develop a business rule that would identify addresses at which to conduct a complete set of in-person NRFU contacts, and
- Using the information for characteristic imputation, particularly relationship to the householder.

7.2. State Program Data

The Census Bureau continues to reach out to states to acquire program participation data for programs such as:

- Supplemental Nutrition Assistance Program (SNAP),
- Women, Infants, and Children (WIC),
- Temporary Assistance for Needy Families (TANF), and
- Alaska Permanent Fund Dividend (PFD).

The Census Bureau currently has signed agreements with over 25 states. The type of data (i.e., the program) varies by states. Some states have agreed to supply SNAP, WIC, and TANF data while other states have agreed to supply a subset of these programs.

Like the CARRA Kidlink file, we are researching possible uses of these state program data sources, including:

- Adding the files to the set of core sources,
- Using the data to develop a business rule that would identify addresses at which to conduct a complete set of in-person NRFU contacts, and
- Using the information for characteristic imputation, particularly race and Hispanic origin.

We are assessing the quality of the data, but we will also need to assess the timeliness of the data being delivered based on the desired use. If we want to use a given state data source in our NRFU operation to help determine how many contacts an address receives, then this use would be more sensitive to when the state provides the information to the Census Bureau. We would be most interested in addresses with people who are participating in the program around Census Day (April 1). For other uses such as assigning characteristics, we could possibly use program participation data from multiple years.

7.3. Third-Party Sources

We are also continuing to research how we can use additional third-party sources to improve our process. One example is the Multiple Listing Service (MLS) data obtained from Corelogic. This third-party file has records of addresses that are on the real estate market. We are assessing the coverage of this file. There are numerous MLS groups across the country, and Corelogic does not have an agreement with all of them. We are researching if this source can help with our determination of vacant addresses.

8. Conclusion

This paper provides details of the AR modeling methodology to determine occupied, vacant, and non-existent addresses that can receive fewer contacts during the NRFU operation. This paper provided results on the use of AR in the 2016 Census Tests. We described how the results from these tests have influenced the evolution of our methodology, ultimately leading to the contact strategy for the 2018 End-to-End Test. This design change uses the UAA information for an

additional mailing sent in May to inform which of these cases will receive the AR treatment and which cases will be reintroduced into the NRFU field workload. This new requirement based on the results of the additional mailing will help guard against misclassification error for the AR determinations.

We also included in this analysis where we nationally set a cutoff that determined 10 percent of NRFU to be administrative record vacant and 15 percent to be administrative record occupied. In order to show potential subnational implications, we included how those determinations were distributed by the percent of Hispanic and the percent of Non-Hispanic Black population in the tract. For the Hispanic concentrations, we see decreasing AR determinations as the concentration increases. For the Non-Hispanic Black concentrations, we see similar results for AR occupied. For the AR vacant units, we see that 12 percent of the addresses in tracts with 50 percent or more concentration of Non-Hispanic Black population would be classified as vacant. This large amount could lead to potential net undercoverage if these units were occupied on Census Day. The proposed 2018 contact strategy helps guard against this by requiring UAA determination on both Census Day and the about 6 weeks later during the start of the NRFU operation.

This paper includes how we are potentially using information from administrative record and third-party sources for the assignment of characteristics. This is being done for cases where we would be using rosters from administrative record sources that did not self-respond to the census. In addition, this paper included information about the usage of the administrative record and third-party information is the characteristic imputation processing for persons that did not report certain characteristics.

The last section includes information on how we are continuing to research additional files. We will be making a final decision on sources to use in September 2018. Researching these additional sources address recommendations from previous working groups about trying to obtain files to address hard-to-count populations.

.

References

Cresce, A. (2012). "Evaluation of Gross Vacancy Rates From the 2010 Census Versus Current Surveys: Early Findings from Comparisons with the 2010 Census and the 2010 ACS 1-Year Estimates," Social, Economic, and Housing Statistics Division Working Paper Number 2012-07. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2012/demo/SEHSD-WP2012-07.pdf

Dowling et al. (2016). Final Report of the National Advisory Committee on Racial, Ethnic and Other Populations Administrative Records, Internet and Hard to Count Population Working Group. Available at: https://www2.census.gov/cac/nac/reports/2016-07-admin_internet-wg-report.pdf

Luque, A. and Wagner, D. (2015). "Assessing Coverage and Quality of the 2007 Prototype Census Kidlink Database." Center for Administrative Records Research and Applications Working Paper #2015-07. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-07.pdf

Rastogi, S. and O'Hara, A. (2012). "2010 Census Match Study," 2010 Census Program for Evaluations and Experiments, No. 247. Available at: https://www.census.gov/2010census/pdf/2010_Census_Match_Study_Report.pdf

Steinhardt et al. (2014). Final Report of the National Advisory Committee Administrative Record and Third Party Data Use in the 2020 Census Working Group. Available at: https://www2.census.gov/cac/nac/reports/2014_2020census_wg.pdf

U.S. Census Bureau (2016). "2020 Census Operational Plan," Version 2.0, U.S. Census Bureau. Available at: https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan2.pdf

Walker, S., Winder, S., Jackson, G., and Heimel, S. (2012). "2010 Census Nonresponse Followup Operations Assessment," 2010 Census Planning Memoranda Series, No. 190. Available at: https://www.census.gov/2010census/pdf/2010_Census_NRFU_Operations_Assessment.pdf