

RESEARCH REPORT SERIES  
(*Statistics #2016-04*)

**Using 2010 Census Coverage Measurement  
Results to Compare Census Nonresponse Followup  
Proxy Responses with Administrative Records**

Mary H. Mulry  
Andrew D. Keller

Center for Statistical Research & Methodology  
Research and Methodology Directorate  
U.S. Census Bureau  
Washington, D.C. 20233

Report Issued: August 30, 2016

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



# Using 2010 Census Coverage Measurement Results to Compare Census Nonresponse Followup Proxy Responses with Administrative Records

Mary H. Mulry and Andrew D. Keller

## 1. Introduction

Currently the U.S. Census Bureau is conducting research on ways to use administrative records to reduce the cost and improve the quality of the 2020 Census Nonresponse Followup (NRFU) at addresses where the Census Bureau did not receive a self-response electronically or by mail. Regardless of the number of contact attempts the 2020 Census NRFU design permits, enumerators will confront the problem of not being able to contact the residents at some addresses. In previous censuses, the strategy at this point has been to find a knowledgeable person, such as a neighbor or apartment manager, who could provide the census information for the residents, called a proxy response. The Census Bureau's recent advances in merging federal and third-party databases to create households that can be used for census enumeration purposes raises the question: Are proxy responses for NRFU addresses more or less accurate than the administrative records available for the housing unit?

Our study attempts to answer this question by comparing the quality of the proxy responses in the 2010 Census with administrative records for the same housing units. The comparison of the quality of the two sources uses the results of the 2010 Census Coverage Measurement (CCM). The goals of our study also include examining whether the quality of proxy responses for NRFU addresses vary by the number of contact attempts prior to the proxy response and/or by whether the administrative records available for the address are deemed high quality or low quality, with defining high quality as part of the research. The evaluation of the quality of the proxy responses and the administrative records files includes comparisons of the number of people enumerated, the number of people correctly enumerated, and the demographic distributions. To provide context, our study also examines the quality of NRFU data from respondents who are household members and the administrative records available for the same addresses.

This report describes the results of the first phase of our assessment. The second phase continues and includes a comparison of demographic characteristics of NRFU proxy responses and ARs in corresponding HUs. Another aspect is to use decision trees in developing statistical models to identify the characteristics of NRFU HUs with corresponding administrative records that have a high probability of being correct. Much of Sections 1, 3, and 4 in this report also can be found in Mulry and Keller (2015).

## 2. Background

The Census Bureau is conducting a series of tests to examine the implementation of adaptive strategies for conducting Nonresponse Followup (NRFU) of the housing units that do not self-respond in a census. The proposed strategies include using administrative records and a variable number of contact attempts with the goal of reducing costs and improving data quality.

An illustration of the strategies under examination comes from the most recent test is the 2014 Census Test, which had a Census Day of July 1. NRFU in 2014 Census Test has four panels that employ

different contact strategies combined with the use administrative record (AR) files in different ways, including a control panel that uses no administrative records. One panel has an adaptive contact strategy while the others have a fixed contact strategy.

- Panel 1 mimics the 2010 Census NRFU contact strategy as closely as possible. A maximum of six contact attempts is allowed with a proxy response permitted only after the maximum attempts to interview a household member have failed. Addresses where interviewers could not obtain a proxy or the proxy responses were not data-defined receive count imputations. The contact strategy is fixed for all households and does not use administrative records in any way. This panel serves as a control for comparing the results of the other panels.
- Panel 2 has a strategy that permits fewer contact attempts than allowed in Panel 1. The Reduced Contact Strategy calls for the first contact attempt to be in person. If the interviewer cannot contact a household member or resolve the address as vacant or not having living quarters (delete) on the first attempt, then the second attempt is by telephone using a number provided by the Census Bureau through a matching the address to a commercial database. If the attempt to contact by telephone is unsuccessful, the interviewer makes a second personal visit. If the interviewer still is unable to contact a household member, then the interviewer takes a proxy. Addresses where interviewers could not obtain a proxy or the proxy responses are not data-defined receive count imputations. This approach uses a fixed contact strategy for all households and does not use administrative records in any way. This panel serves to evaluate the Reduced Contact Strategy without any confounding from the use of administrative records.
- Panel 3 employs administrative records as much as possible. First, addresses identified by administrative records as not having living quarters (delete), vacant, or having administrative records of sufficient quality to use for enumeration are removed from the NRFU field workload. Then Panel 3 used the fixed Reduced Contact Strategy described for Panel 2 for all remaining addresses. As in Panel 2, count imputations are created for addresses where interviewers could not obtain a proxy or the proxy responses are not data-defined. This panel serves to evaluate the cost and operational aspects of extensive use of administrative records.
- Panel 4, uses administrative records at several points to reduce the field workload and employs an adaptive design in the field implementation. This panel requires developing a two-level assessment of the quality of the data for the address in the administrative records file, high quality and low quality. The approach first uses an AR file to determine the occupancy status of each housing unit (HU) that did not mail back a census questionnaire. The HUs that ARs indicate are vacant receives the designation of vacant, and addresses that ARs show do not have living quarters are classified as deletes. Then interviewers will attempt to contact only those HUs that ARs indicate are occupied.

Each NRFU housing unit in Panel 4 will receive at least one contact attempt. The first attempt will be by telephone using centralized CATI if a number can be found and in person if a number is not available. If the interviewer is able to obtain an interview on the first attempt, then the HU has a resolved status. The HUs that interviewers are not able to contact on the first attempt continue to have the status of unresolved and are divided into two groups, those that have high quality ARs and those who do not. The interviewer's handheld device will be loaded with a designation of which HUs have high quality ARs and which do not. When the HU has high quality ARs, the case is removed from his/her workload, and AR information is used to determine the population in the HU. When the HU does not have high quality ARs, the interviewer

continues to attempt to obtain information about residents of the HU and may take a proxy interview the same day as the first contact attempt. If the interviewer does not obtain an interview with a household member or a proxy on the first day, the interviewer may return to attempt to get a proxy for a maximum of three attempts over the two days. If the interviewer is able to get a proxy on the second day, that information determines the population of the HU. If the interviewer is unable to get a proxy on the second day, the population of the HU is determined by inserting AR information, or if there is no AR information for the HU, then a count imputation is made.

The results of the four panels provide information for the design of the 2015 Census Test. For all panels in the 2014 Census Test and the panels designed for subsequent tests, the question arises as to whether the proxy responses are more accurate than ARs available for the NRFU HUs. Even though the ARs are not deemed of the highest quality, the ARs could be of better quality than the proxy responses. Fieldwork is expensive. If the proxy interviews are not of higher quality than ARs, the additional cost of the fieldwork may not be justified.

Ideally, the 2014 Census Test NRFU proxy response for a HU and the ARs for the HU could be compared against a ‘gold standard’ interview conducted by a highly skilled interviewer with the residents of the HU. Then a determination could be made as whether the proxy or the ARs had better information, or whether they were of comparable quality. However, the 2020 Census testing cycle has a tight timeframe that does not allow for a ‘gold standard’ interview operation.

Instead, the plan is to compare the quality of the 2010 Census NRFU HUs with proxy responses and the AR data for those HUs using the results of the 2010 Census Coverage Measurement (CCM) in a sample of block clusters. The approach is similar to a methodology discussed in Mulry and Spencer (2012).

### **3. Research Approach**

#### **3.1 Research questions**

The focus of our research is to answer the following questions to produce information useful for the design of the strategy for contacting HUs during the 2020 Census NRFU:

- Are proxy responses for NRFU addresses more accurate than the administrative records available for the housing unit or are they less accurate?
- Does the quality of proxy responses for NRFU addresses vary by the number of contact attempts prior to the proxy response and/or by whether the administrative records available for the address are deemed high quality or low quality?

#### **3.2 Population**

The population under study is defined as the people whose Census Day residence is in a housing unit enumerated in the 2010 Census NRFU by a proxy respondent, and administrative records are available for the housing unit. According to Census residency rules, the correct address for a person’s enumeration is his/her usual residence around Census Day, which is April 1 of the census year. We consider the quality of two lists of the population using the criteria of whether the person is found on the list at the correct location on Census Day according to Census residency rules. One list of this population is the census enumerations, and the other list is the administrative records for the same housing units.

For context, we also examine the quality of NRFU enumerations where the respondent is a household (HH) member and the administrative records at these addresses.

In this study, the definitions of the populations enumerated by proxy and HH member respondents are operational and depend on the conduct of the 2010 Census operations. The HUs enumerated by HH member respondents failed to self-respond by mail. The HUs enumerated by proxy failed to self-respond by mail, and none of the HH members gave an interview to an NRFU enumerator. In 2010, enumerators had to make six contact attempts prior to taking a proxy interview. Therefore, our analyses, as well as the population definition, are conditional on the type of response observed in the 2010 Census. In addition, the analysis is conditional on the sources of administrative records that we consider.

### **3.3 Gold standard**

The assessment of the quality of the proxy responses and the records in the selected administrative files takes advantage of the extensive fieldwork, processing, and clerical matching conducted for the CCM, which is the justification for using the CCM results as a ‘gold standard.’ The 2010 CCM was designed to measure census coverage error with a post-enumeration survey composed of two samples, the enumeration sample (E sample) and the population sample (P sample). The E sample and the P sample used the same sample of block clusters. The E sample contained the census enumerations in the block clusters and its design supported the estimation of erroneous enumerations. The P sample constructed its list of the population in the block clusters independently of the census and was designed to support the estimation of census omissions. Each P-sample and E-sample record that CCM processed was assigned a residence code indicating one of the following: (1) the person was a resident of the sample block cluster on Census Day, (2) was not a resident on Census Day, or (3) had unresolved Census Day residence.

The P sample interviews occurred in August and September 2010 independently from the 2010 Census. These interviews collected data that enabled constructing the Census Day (April 1) roster for the address by asking when current residents moved to the address and about any Census Day residents who had moved from the address. The Census Bureau used a combination of electronic and clerical operations to match the P-sample people to the 2010 Census enumerations and conducted follow-up interviews in February 2011 to collect additional data when a person’s Census Day residence could not be resolved. The CCM operation determined whether the census enumerations and P-sample persons were residents of their sample block cluster or the blocks surrounding the block cluster on Census Day assigning the statuses of resident, nonresident, and unresolved. The CCM built this tolerance to avoid including minor geocoding error or mail delivery mistakes in the coverage error estimates, which would increase the variability of the estimates.

Since the P sample is available only for the block clusters in the CCM sample, the comparison has to be restricted to the CCM block clusters. Although the 2010 CCM estimation does not require assuming that the P-sample interview is the ‘truth,’ the P-sample interviews are believed to be of higher quality because the interviewers have more training and experience since they were chosen from the pool of the best NRFU interviewers. In addition, the CCM interviewers were supported with a Computer Assisted Personal Interviewing (CAPI) instrument and given additional residence probes to ask.

The NRFU enumerations in the E sample have residence status codes assigned during the CCM processing, but the administrative records in the NRFU HUs do not. We link the administrative records to the E and P sample records to retrieve CCM residence status codes. When a person’s administrative record links to an enumeration in HU enumerated by a proxy response at the same address, the CCM residence code for the proxy response will indicate whether the person’s enumeration at the address was correct. For example, if the person was enumerated at two addresses and the address not in the sample block was the correct Census Day residence, the enumeration in the sample block cluster was coded erroneous. This would mean the location of the person’s administrative record was also in error.

However, when a proxy response for a person and the administrative record file disagree, the CCM results provide information about whether the person should have been enumerated at the address and therefore, whether one of the sources is better for the person. Requiring the same address for a person's administrative record and the linking NRFU enumeration to retrieve a CCM residence code lends credibility to the assumption that the person lived at or is associated with the address. An administrative record will be inserted in the census at its address if the Census Bureau decides to use administrative records as enumerations. Requiring the same address from both sources means the correct enumeration rate reflects the accuracy of the use of administrative records at the addresses where they will be inserted in the census.

### 3.3 Underlying assumptions

This study approach has five major underlying assumptions:

- The results for proxy interviews in NRFU in the 2010 Census are applicable to the proxy interviews that would occur in the 2020 Census. The implementation of self-response and NRFU in the 2020 Census will be different from what occurred in the 2010 Census, and in particular, the procedures for taking proxy interviews in NRFU will differ.
- The 2010 CCM was able to determine whether the people on the rosters in NRFU proxy interviews were enumerated at the correct location, meaning their usual residence.
- The electronic matching algorithm used in this study (described in Section 3.5) was able to link a person's administrative record to the same person's record in the CCM P and E samples.
- The records from the administrative sources used in this study reflects the future availability from these sources.
- When a person's address in administrative records matches their address in NRFU, the person lives at or is associated with the address.

### 3.4 Data

For this study, we are going to focus on HUs in the CCM sample block clusters that were on the NRFU list in the E sample and on the independent list of HUs created for the P-sample, and call this group the *combined CCM*. We need both E-sample and P-sample records because some or all the records for an occupied HU on the census list may be whole person imputations, but the P-sample interviewers were able to obtain data for the residents. In addition, the P sample may have information regarding persons in ARs not listed on the census form. We use the combined CCM to look up residence status codes for the administrative records. We do not form estimates using the combined CCM.

The administrative records file is the merger of two files unduplicated within housing units: (1) the IRS 1040 forms filed in all months of 2010, (2) the Medicare records for all months of 2010. One reason the files were not unduplicated across housing units is that when duplicate records appear, there is no way to determine which is at the person's usual residence on Census Day. The files from these two sources contain data for whole households. In addition, the 2014 Census Test operations used only these two sources.

The combined CCM contains 27,724 HUs that were proxy responses in NRFU with 10,416 occupied in NRFU, 15,012 vacant and 2,296 deleted because they did not have living quarters. Table 1 shows that of the 10,416 occupied housing units, 5,310 also have administrative records, the implication being that 5,106 have no records in the AR files we are using. For comparison, the percentage of the 144,000 occupied HUs in NRFU that have records in the combination of IRS 1040 and Medicare files is 56%, which means the combined CCM percentages are reasonable with proxy HUs being a little lower at 51% and the HH member HUs being a little higher at 61.3%.

**Table 1.** 2010 Census NUFU HUs in the combined CCM by AR status and type of NRFU respondent (unweighted)

AR status of HUs	Proxy		HH Member	
	HUs	%	HUs	%
Person records on AR list	5,310	51.0%	16,876	61.3%
No person records in AR list	5,106	49.0%	10,647	38.7%
Total	10,416	100.0%	27,523	100.0%

Note: ARs include IRS 1040 forms and Medicare records for all of 2010.

For the NRFU HUs in Table 1 that have AR records, Table 2 shows the distribution of the number of NRFU person records enumerated by proxy and HH member respondents and the corresponding number of records for the same HUs. In each of the two sources, the size of population in the proxy HUs is about 25% of the size of population in the HH member HUs. The AR file has more people in HUs enumerated by proxy than NRFU but fewer people in the HUs enumerated by HH members. Combining all the NRFU HUs, the AR file has 505 records more than NRFU, about a 0.8% difference. Late in the analysis we discovered that 88 of the AR persons in the proxy HUs and 237 in the HH member HUs had died in 2009. These remain in the analysis but we do address this issue for AR file construction in the recommendations in Section 7.

**Table 2.** Number of individual records found in AR files and number of individual records found on the combined CCM list in HUs in the combined CCM and occupied in the census by type of NRFU respondent.

Respondent type	ARs	NRFU
Proxy	12,880	11,766
HH member	50,876	51,485
Total	63,756	63,251

The 5,310 HUs with ARs had 11,766 NRFU enumerations of persons with 9,258 that had at least two characteristics, one of which could be a name, which was considered enough information to be an enumeration, called *data-defined*. The remaining 2,508 were whole person imputations. Therefore, the imputation rate in these HUs is 21.3%, which is lower than the national average of 23.1% for imputations among NRFU proxy enumerations.

For completeness, we note that our analysis does not include 1,048 HUs with proxy respondents in the E sample that are not also on the P-sample list, making them ineligible for the combined CCM list. The number of these HUs containing ARs is 231 resulting in 460 ARs for persons not being evaluated. In addition, the study does not include the 6,154 HUs on the P sample list that were not on the E sample list.

### 3.5 Matching ARs to combined CCM

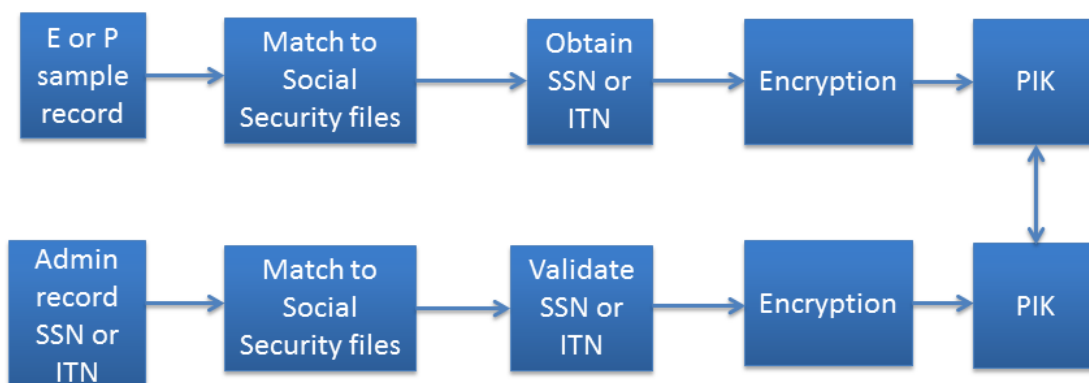
The comparison of the 2010 Census NRFU HUs with proxy responses and the AR data for the HUs in the CCM block clusters requires linking the AR records to the combined CCM to retrieve residence codes assigned during the CCM processing. The linking between the AR data and the combined CCM requires that both sources have Protected Identification Keys (PIKs), which are essentially encrypted Social Security Numbers (SSNs) or Individual Tax Identification Numbers (ITNs, included when we use the abbreviation SSN). AR data comes with SSNs that the Census Bureau staff converts to PIKs after a validation of their accuracy through matching to Social Security Administration (SSA) files, a procedure called the Person Identification Validation System (PVS) (Wagner and Layne 2014). When a data file with records for persons does not come with SSNs, the Census Bureau uses its system to look up SSNs in



SSA files and encrypt them by assigning PIKs. PIKs have been assigned to the 2010 Census so the NRFU enumerations in the HUs with proxy responses have PIKs. PIKs also have been assigned to all the names collected in the P-sample regardless of the ultimate classification of nonmover, in-mover, out-mover, or never a resident of the sample block. . Figure 1 illustrates the process of assigning PIKs and linking the files.

**Figure 1.**

## PVS assigns Protected Identification Keys (PIKs) for matching between CCM E & P sample records and admin records



Having the CCM results available to compare the proxy responses and AR records is important because the estimated correct enumeration rate for the 2010 Census was 70.1% for persons enumerated by proxy respondents with 23.1% having all characteristics imputed, 5.6% being duplicates, and 1.1% being erroneous for other reasons. In contrast, 93.4% of the persons enumerated by a household member in NRFU were correct with 1.6% having all characteristics imputed, 4.2% being duplicates, and 0.8% being erroneous for other reasons (Mule 2012, Keller and Fox 2012). Even though enumerations that had all characteristics imputed, called *whole person imputations*, were not processed in the CCM E-sample due to lack of information to identify a person uniquely, the corresponding HU was included in the CCM P-sample and will usually have information about the residents that can be used for evaluating an AR records associated with the address. The P sample also may have residency information for enumerations that were data-defined but had insufficient information to be processed in the CCM. The CCM requirement for sufficient information was a name and at least two characteristics because the CCM operations matched the enumerations to the names on the P-sample interview rosters.

When a person was enumerated by a proxy response and in the AR file at the same address, the CCM residence code for the proxy response indicates whether the person's enumeration at the address was correct. If a person appears in the AR file but does not link to a combined CCM record at the same address, we can search the PIKs assigned to 2010 Census enumerations to learn if the person was enumerated elsewhere, but are not able to assess the accuracy for enumerations outside the CCM sample block clusters. If the person has an enumeration elsewhere that could not be assigned a PIK, we are not able to detect it using PIK matching.

Other types of electronic matching algorithms that do not rely on the assignment of PIKs, such as the household-based matching used by CCM, were not attempted. Household-based matching may or may not identify additional links between ARs and the combined CCM. Regardless, our results must be viewed as conditional on the use of PIK matching.

### 3.6 Evaluation criteria

The evaluation of the quality of enumerations from the proxy responses and records in the AR file in the same HUs includes the rate of correct enumerations. The assessment also includes comparing the count of persons in each source. Comparable calculations are made for enumerations and AR records in HUs with HH member responses.

- The total number of people enumerated at the sample addresses in each source
- The total number of people correctly enumerated at the sample addresses in each source.
- HUs classified by (1) all ARs are at the correct Census Day residence, (2) at least one AR is erroneous (not at the Census Day address) or its Census Day residence is unresolved, and (3) at least one Census Day resident does not have an AR at the address.
- The number of people who are less than 18 years of age versus the number who are 18 years of age or over whose age was reported or imputed in NRFU by whether their enumeration received a PIK and was in the correct location.
- The number of people who are Hispanic versus the number nonHispanic whose Hispanic ethnicity is reported and imputed in NRFU by whether their enumeration received a PIK and was in the correct location.

## 4. Results

Although the focus of our analyses is the NRFU HUs enumerated by proxy respondents, we are going to present results for NRFU HUs enumerated by household (HH) members for comparison. First, Section 4.1 considers the quality of the records for persons under the criteria of whether the address on the record is in the correct location as determined by CCM. Analyzing the quality of individual records provides insight when viewing the quality of the records for complete households, which is the focus of Section 4.2. In addition, analyses of individual records provide information about several potential uses of administrative records, such as for enumeration and for use in developing imputation models.

### 4.1 Analysis of individual person records

Even though Table 2 shows the number of records in ARs and NRFU generally agree, this alone is not enough to evaluate the quality of the individual records in the two systems. We need to know whether a person's record is at the correct location of the person's Census Day residence and whether the characteristics of the person and the size and composition of the households are correct.

Two things have to happen to evaluate an AR for a person: (1) the AR PIK has to link to a record in the combined CCM, (2) the combined CCM record has to have a resolved residence status.

Table 3 shows the weighted distribution of combined CCM residence status for enumerations and ARs in NRFU HUs in the combined CCM by NRFU respondent type while Table 3U in the Appendix shows the same results unweighted. The first thing to notice is that the unweighted and weighted distributions of

CCM residence status is very similar for each NRFU respondent type. The weighted and unweighted distributions for the ARs in HUs by NRFU respondent type also are similar. The weights are the CCM E-sample block cluster weights not adjusted for CCM nonresponse. Since the CCM sample design was able to keep the block cluster weights within a tight range, the similarity of the unweighted and weighted distributions is reasonable. We use the weighted results in our discussion.

To compare the distributions of the residence statuses from different types of respondents or different sources, we perform a chi-square test using the Rao-Scott adjustment (Lohr 1999) to account for the sampling design. For the design effect of the CCM sample, we examined Table 8 in Olson and Griffin (2012) that contains the means of several ranges of the observed correct enumeration rate, the number of observations in each range, and the standard error of the mean. The design effects varied between 2.5 and 3.5 across the categories. We use a design effect equal to 3 for the Rao-Scott adjustment to the chi-square statistics. For the chi-square tests, we use four cells: correct residence, erroneous residence, unresolved residence, and unable to process. For NRFU, we define the unable to process cell by collapsing insufficient information for CCM and whole person imputations, and for ARs, we collapse the records found at another census address and those not linked to a census record.

For the NRFU proxy enumerations, Table 3 shows that CCM found that 56.6% were at the correct residence, and 4.1% were at an erroneous residence. CCM attempted but could not determine Census Day residence for 15.8% of the NRFU proxy enumerations. CCM did not attempt to process the 2.8% that had insufficient information or the 20.7% that were whole person imputations.

**Table 3. Weighted distributions of combined CCM residence status for enumerations and ARs in NRFU HUs in the combined CCM by NRFU respondent type (shown in 1,000's)**

Census Day residence status	Proxy respondent			
	NRFU		AR	
	count	%	count	%
<b>Correct residence</b>	5,235.2	56.6%	5,017	49.1%
<b>Erroneous residence</b>	380.9	4.1%	418	4.1%
<b>Unresolved residence</b>	1,462.4	15.8%	379	3.7%
<b>NRFU not processed by CCM</b>				
<b>Insufficient info</b>	258	2.8%	-	-
<b>Whole person Imputation</b>	1,903	20.7%	-	-
<b>AR PIK not in census at same address</b>			4,397	43.1%
<b>Total</b>	9,257	100.0%	10,212	100.0%

Census Day residence status	HH member respondent			
	NRFU		AR	
	count	%	count	%
<b>Correct residence</b>	36,720.2	88.0%	29,971	72.5%
<b>Erroneous residence</b>	1,058.9	2.5%	1,054	2.5%
<b>Unresolved residence</b>	2,308.2	5.5%	1,283	3.1%
<b>NRFU not processed by CCM</b>				
<b>Insufficient info</b>	258	2.6%	-	-
<b>Whole person Imputation</b>	1,903	1.4%	-	-
<b>AR PIK not in census at same address</b>			9,038	21.9%

<b>Total</b>	41,741	100.0%	41,346	100.0%
--------------	--------	--------	--------	--------

For the NRFU enumerations from HH members in Table 3, we see that 88.0% are at the correct residence, 2.5% are at an erroneous residence, and 5.5% had an unresolved residence status. However, 2.6% had insufficient information for CCM to process and 1.4% of the proxy enumerations were whole person imputations, which CCM did not process.

A chi-square test comparing the distributions of the residence status of the NRFU enumerations for the two types of respondents produced a p-value less than 0.001, and therefore, we conclude that the distributions are different. We see that the percentage of proxy enumerations that are at the correct residence 56.6% is lower than the percentage of HH member enumerations at 88.0%. The most apparent difference is that the percentage of whole person imputations is much higher for the proxy enumerations at 20.7% than for the HH member respondents at 1.4%. However, the HUs that are remaining after the attempts to get HH member respondents fail get rolled over to the attempts to get proxies so virtually all the whole person imputations get attributed to the proxies, although both the self-response phase and the NRFU HH member response phase also failed to get a response.

Turning to the residence status of the ARs in NRFU HUs with proxy respondents in Table 3, links to combined CCM records showed that 49.1% were at the correct residence, 4.1% were at an erroneous residence, and 3.7% had an unresolved residence. The percentage that did not link at the same address and could not be evaluated is 43.1%. For some insight about the ARs that did not link, the unweighted data in Table 3U in the Appendix shows that 17.3% of the proxy responses did not link to a combined CCM record at the same address but linked to an enumerations elsewhere in the census while 26.8% of the proxy responses did not link to a combined CCM record at the same address or elsewhere in the census.

When we examine the ARs in the HUs with HH member respondents, we see that links to the combined CCM found that 72.5% were at the correct residence, 2.5% were at an erroneous residence, and the residence status of 3.1% could not be resolved. The percentage that did not link at the same address and could not be evaluated is 29.1%. The unweighted data in Table 3W in the Appendix shows that 10.5% did not link to a combined CCM record at their AR addresses but were found at other addresses in the census while 12.9% did not link to the combined CCM at their AR addresses or at another address in the census.

The chi-square test to compare the distributions of the ARs for the two respondent types produces a p-value of 0.010, which indicates that the distributions are different. The percentage of ARs that are at the correct residence is 49.1% in the HUs enumerated by proxy while the percentage correct is higher at 72.5% in the HUs enumerated by a HH member. There is not much difference in the percentages of the ARs that at an erroneous address or with an unresolved residence. However, the percentage that did not link at the same address and could not be evaluated is higher for proxy respondents 43.1% than for HH member respondents at 21.9%.

Next, we compare the distributions of the residence statuses for the NRFU enumerations and the ARs by respondent. For the HUs with proxy respondents, the chi-square test produced a p-value less than 0.001, which leads us to conclude that the distribution of the residence statuses for the NRFU enumerations and the ARs in these HUs are different. For the HUs with HH member respondents, the p-value of the chi-square test is 0.028, which indicates the distributions of the residence codes are different. For both types of respondents, the percentage of NRFU enumerations at the correct residence is higher than observed for ARs, and the percentage of ARs that cannot be evaluated is higher than observed for NRFU enumerations.

Both NRFU and ARs have a substantial percentage of records where this approach is unable to evaluate their residence status. The seemingly high percentage of records that do not link to a combined CCM record at their AR address but link to a census address elsewhere causes concern that these ARs are not at

the correct Census Day residence and more importantly, that inserting them as census enumerations would create duplicate enumerations. Since the CCM sample did not include the address where AR PIKs were found, the CCM did not evaluate accuracy of the enumeration of the people at the address. Therefore, the accuracy of AR records that linked to these enumerations also could not be evaluated.

Interestingly, the percentage of records with a CCM resolved residence status is higher for NRFU enumerations than ARs in both HUs with both types of respondents. Keep in mind that all the AR records have PIKs, but the Census Bureau procedure may or may not be able to assign PIKs to the census enumerations.

The assignment of PIKs to the combined CCM records proved crucial to being able to evaluate the ARs in HUs enumerated during NRFU. Therefore, the percentage of NRFU enumerations that received PIKs is an evaluation tool in and of itself. Table 4 shows the distribution of the residence status of enumerations with PIKs and those without PIKs by NRFU respondent. Of the NRFU enumerations where the PVS system attempted to assign PIKs, 73% (SE = 0.9%) of those in HUs enumerated by proxy received PIKs while 92% (SE = 0.2%) of those enumerated by a HH member received PIKs. If the whole person imputations are included, the percentage is 58% (SE=0.8%) for proxy respondents and 91% (SE=0.2%) for HH member respondents. When whole person imputations are included and when they are not, the tests of difference between the percentages of enumerations assigned PIKs for proxy and HH member respondents produced p-values less than 0.001 so we conclude there is a difference in the enumerations from the two types of respondents.

**Table 4. Weighted** distributions of combined CCM residence status for enumerations in NRFU HUs by NRFU respondent type and PIK status (shown in 1,000's)

Census Day residence status	Proxy			HH member		
	with PIK	without PIK	Total	with PIK	without PIK	Total
<b>PIK attempted</b>						
Correct residence	3,625.8	1,609.4	5,235.2	34,322.1	2,398.2	36,720.2
Erroneous residence	266.4	114.5	380.9	844.0	214.9	1,058.9
Unresolved residence	337.5	173.2	510.7	1,713.5	594.7	2,308.2
Insufficient info for CCM	1,124.9	85.1	1,210.0	990.8	80.1	1,070.9
<b>Subtotal</b>	5,354.6	1,982.1	7,336.7	37,870.3	3,287.9	41,158.3
	73%	27%	100%	92%	8%	100%
<b>PIK not attempted</b>						
Whole person imputation		1,920.6	1,920.6		583.0	583.0
<b>Total</b>	5,354.6	3,902.8	9,257.4	37,870.3	3,870.9	41,741.2
	58%	42%	100%	91%	9%	100%

In summary, a distinguishing feature that indicates the quality of NRFU enumerations appears to be whether they can be assigned a PIK. Those that receive PIKs tend to be in the correct location at high rate. Table 5 shows the correct enumeration rate for several criteria for the denominator for enumerations with and without PIKs by type of NRFU respondent. We do not conduct statistical testing but use the data in Table 5 to illustrate the effect of the choice of the denominator of the correct enumeration rate.

When the denominator includes only the enumerations where CCM could resolve the residence status, namely those correct and erroneous, the percentage correct is not dramatically different from the percentages for the HH member respondents without PIKs and both categories for proxy respondents, which range from 92% to 98%. By the way, from Table 3 the percentage of AR records with a resolved residence status in proxy HUs that are correct, which is 92% ( $5,017/(5,017+418)$ ), is in the same range.

For the data-defined enumerations with PIKs, 68% from proxy respondents and 91% from HH member respondents are in the correct location. However, the correct enumeration rate among enumerations that are data-defined but not assigned a PIK is 81% for proxy respondents and 73% for HH member respondents. When the denominator for those without PIKs includes whole person imputations, the correct enumeration rate for proxy respondents is 41%. For HH member respondents, rate becomes 62% with the inclusion of the imputations. Keep in mind that whole person imputations are a much smaller percentage of the enumerations by HH members than for proxy respondents.

**Table 5. Weighted** correct enumeration (CE) rate for enumerations in occupied HUs in the combined CCM with several criteria for the enumerations included in the denominator by type of NRFU respondent.  
(shown in 1,000's)

Status of enumerations in denominator	Proxy respondent			HH member respondent		
	Total	CE	% CE	Total	CE	%CE
<b>With PIK</b>						
CCM resolved status	3,892	3,626	93%	35,166	34,322	98%
Data-defined	5,355	3,626	68%	37,870	34,322	91%
<b>Without PIK</b>						
CCM resolved status	1,724	1,609	93%	2,613	2,398	92%
Data-defined	1,982	1,609	81%	3,288	2,398	73%
Data-defined & imputed	3,903	1,609	41%	3,871	2,398	62%

#### 4.2 Analysis of records for entire households

Our ultimate interest is the quality of ARs on a household basis. Our analysis examines two measures. One is the percentage of HUs where the population counts from NRFU and ARs are equal. The other is the percentage of NRFU HUs where the combined CCM determines the AR roster is perfect. These are descriptive analyses with unweighted data.

Table 6 shows that the percentage HUs where the NRFU and AR population counts are the same is 51% for both proxy and HH member respondents. However, the AR population count being equal to the NRFU population count does not mean that the AR roster for the HU has the correct Census Day residents. CCM provides a means to determine the accuracy of the AR roster.

**Table 6.** Unweighted comparison of HU population counts from NRFU and ARs by respondent type

HU population counts	proxy		HH member	
	HUs	%	HUs	%
Same AR & census	2,685	51%	8,633	51%
Different AR & census	2,625	49%	8,243	49%
<b>Total</b>	<b>5,310</b>	<b>100%</b>	<b>16,876</b>	<b>100%</b>

Therefore, we examine the accuracy of the ARs on a household basis for the 5,310 proxy HUs and 16,876 HH member HUs that have ARs. Table 7 shows the percentage of HUs in the following categories as determined by the combined CCM:

- AR Perfect – All AR persons in the HU are Census Day residents at the address and no Census Day residents are omitted from the AR roster.
- AR Erroneous Enumerations and Unresolved Enumerations (E&Us) –At least one AR record in the HU either linked to a combined CCM record coded as not being a Census Day resident at the address or did not link to a combined CCM record with a resolved residence status.
- AR Omissions – There is at least one person that the combined CCM found to be a Census Day resident at the address, but the person(s) is(are) not on AR roster for the address.

**Table 7.** Status of AR records in NRFU HUs in the combined CCM by NRFU respondent type

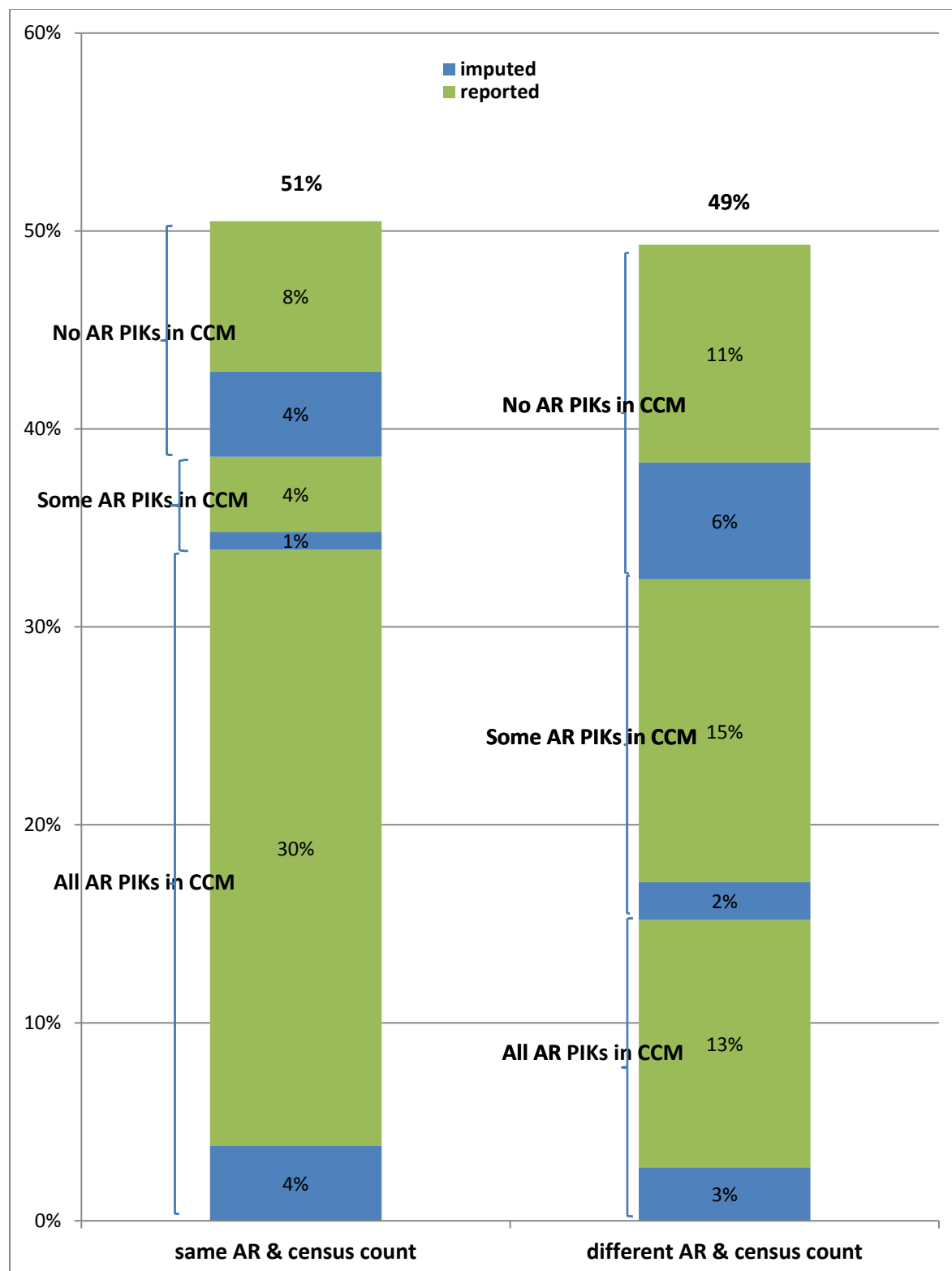
HU status	Proxy		HH member	
AR E&U	3,180	59.9%	6,846	40.6%
AR Perfect	1,722	32.4%	7,256	43.0%
AR Omissions	408	7.7%	2,774	16.4%
<b>Total</b>	<b>5,310</b>	<b>100.0%</b>	<b>16,876</b>	<b>100.0%</b>

When the ARs in the 5,310 proxy HUs are considered on a household basis instead of a individual basis, 1,722 (32.4%) are perfect in that the combined CCM indicated every record as being at the person's Census Day residence and no persons were omitted. We also find that ARs for 408 (7.7%) of the HUs omit at least one person that the combined CCM found to be a Census Day resident at the address. The remaining 3,180 (59.9%) have at least one record that the combined CCM found not to be a resident at the address on Census Day, or the person's Census Day residence was not determined because the AR did not link to a combined CCM record with a resolved residence status.

Surprisingly, the percentage of HUs with HH member respondents who omitted at least one Census Day resident from the AR roster was 16.4%. In addition, 43.0% of the AR rosters for HUs enumerated by HH members are perfect. The percentage of HUs with an AR for at least one person who was not a Census Day resident or had an unresolved Census Day residence was 40.6%.

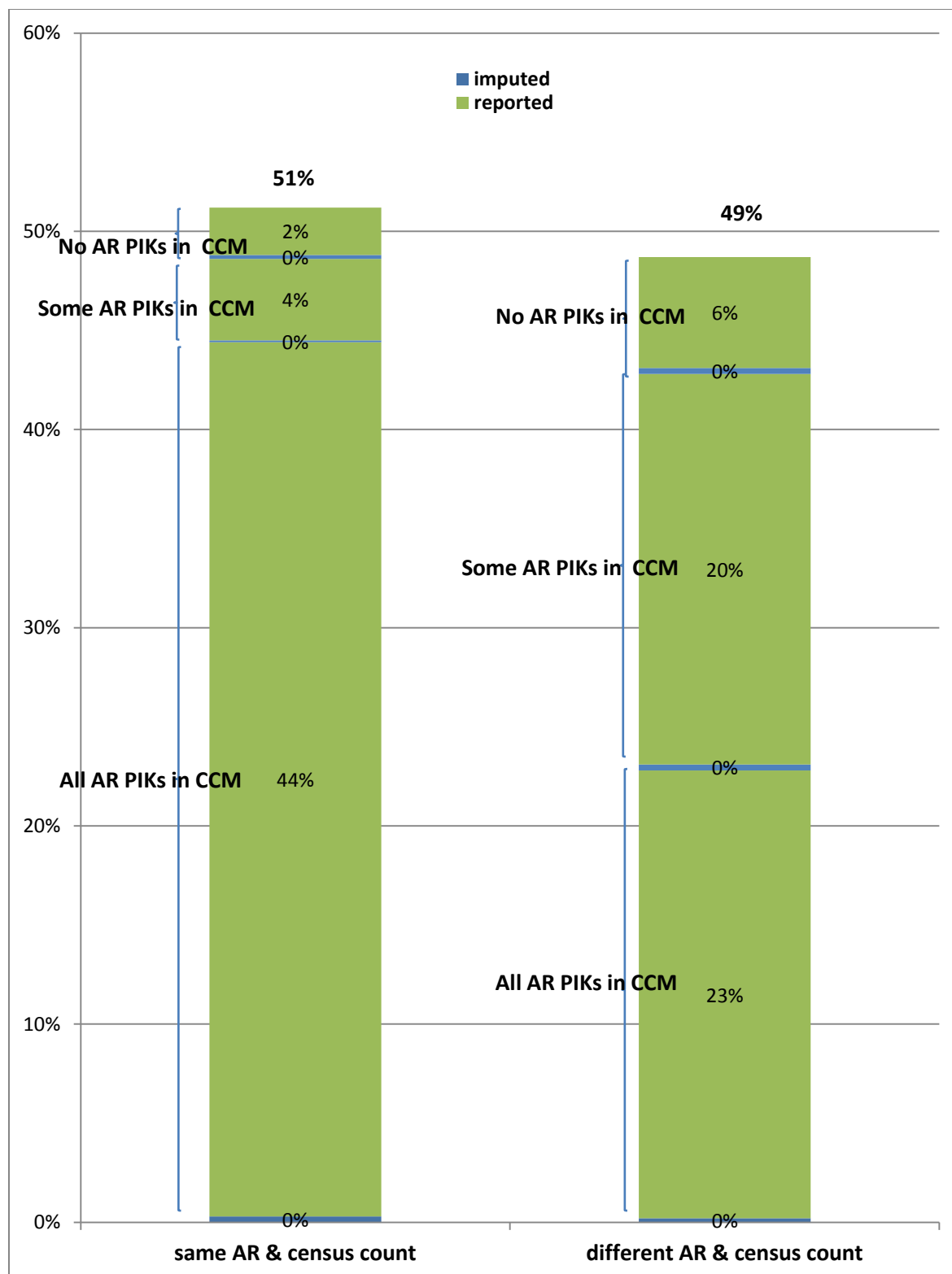
In Figure 2, visual presentations of the distributions of the proxy HUs in Table 6 by whether or not the census and AR counts agree for the HU. Table 6 classifies AR HUs by whether all AR PIKs link to combined CCM record PIKs, some link, or none link. Even though AR records cannot link to census enumerations that are whole person imputations, the HUs often have P-sample records with PIKs available for linking to ARs and providing evaluative information. Figure 2 excludes HUs with a mixture of reported and imputed records since their number is very small. Essentially, either all the enumerations in a NRFU HU are reported or all the enumerations are whole person imputations. Figure 3 shows the same distributions for NRFU HUs with HH member respondents.

**Figure 2.** Distribution of combined CCM status of ARs in NRFU **proxy** respondent HUs by whether the AR HU count agrees with the census count and whether census enumerations in the HU were reported or imputed. Very few HUs have both reported and imputed enumerations. (unweighted)





**Figure 3.** Distribution of combined CCM status of ARs in NRFU **HH member** respondent HUs by whether the AR HU count agrees with the census count and whether census enumerations in the HU were reported or imputed. Very few HUs have both reported and imputed enumerations. (unweighted)



Figures 2 and 3 provide an interesting perspective that enables comparing and contrasting ARs and census enumerations in NRFU HUs with proxy and HH member respondents. The unweighted percentage of proxy HUs and the HH member HUs that have agreement between the NRFU and AR counts is equal at 51%. However, about 13% of proxies HUs have agreement between AR and NRFU counts but all their census enumerations are imputed although 4% have ARs that can be evaluated because all the AR PIKs in these HUs are found in the combined CCM. In contrast, about 1% of the HH member HUs have agreement between the AR and NRFU counts and all their census enumerations imputed. All the imputed categories for HH member respondents show 0% since each is 0.3% or less.

About 16% (13% + 3%) of the proxy HUs have a disagreement between the AR and NRFU count and have all their PIKs in the combined CCM, which means they can be evaluated if the linking CCM record has a resolved code. On the other hand, about 23% of the HH member HUs have a disagreement between the AR and NRFU count and have all their PIKs in the combined CCM so they also may be evaluated.

## 5. Availability of characteristics

An advantage that field data collection should have over administrative records is that that the information is current while the desired information in administrative records files may be out of date or not available. Therefore, an important measure of the quality is whether NRFU responses include demographic characteristics used in important applications of census data. Proxy respondents may tend not to provide demographic information either because they do know it or they are uncomfortable providing it. Also of interest is whether NRFU responses that the PVS assigned PIKs because enumerations assigned PIKs tend to be in the correct location at a higher rate than those without PIKs. In addition, those assigned PIKs possibly can be linked to administrative records that contain the missing demographic characteristics. This section is descriptive in nature and does not include statistical tests.

For an initial indication of the quality of reporting demographic characteristics, we analyze the reporting patterns of Hispanic ethnicity and age, which we restrict to two categories, 0 to 17 years and under 18 years or over. We select the age variable because age is an important characteristic used by the PVS system to assign a PIK to a record. Birthdate is even more helpful than age, particularly for common names. Both age and Hispanic ethnicity are important for drawing legislative districts and managing school districts. Other demographic characteristics also are important and may warrant further analyses.

For Hispanic ethnicity, Table 8 contains the weighted distributions of reported and imputed values for proxy and HH members by the values Hispanic and not Hispanic. Table 8 shows that 69.0% of the enumerations from proxy respondents reported a value while the percentage of enumerations with reported values from HH member respondents is larger at 95.9%.

Another consideration is whether the percentage of Hispanic and nonHispanic values that were reported is comparable within proxy and HH member respondents. Table 8 shows that for the proxy responses, 65.1% of the Hispanic values were reported while 69.8% of the non-Hispanic values were reported. For HH member responses, 64.7% of the Hispanic values were reported and 96.2% of the nonHispanic values were reported. When viewing the distributions within respondent types, keep in mind that the imputation methodology was geographically based and using the entire census, not just NRFU enumerations.

**Table 8.** Weighted distributions of reported and imputed values of the Hispanic variable by whether the respondent was a proxy or a HH member (shown in 1,000s)

	<b>reported</b>	<b>imputed</b>	<b>row total</b>	<b>column %</b>
<b>Proxy</b>				
Hispanic	1,095.5	586.5	1,682.0	19.3%
row %	65.1%	34.9%	100.0%	
not Hispanic	5,288.2	2,287.2	7,575.4	80.7%
row %	69.8%	30.2%	100.0%	
column total	6,383.7	2,873.6	9,257.4	100.0%
row %	69.0%	31.0%	100.0%	
<b>HH member</b>				
Hispanic	8,427.3	476.3	8,904	24.2%
row %	94.7%	5.3%	100.0%	
not Hispanic	31,619.5	1,218.1	32,837.6	75.8%
row %	96.3%	3.7%	100.0%	
column total	40,046.9	1,694.4	41,741.2	100.0%
row %	95.9%	4.1%	100.0%	

Table 9 shows the weighted distribution of the combined CCM residence status of proxy enumerations with reported and imputed values of Hispanic ethnicity. Table 10 shows the same distribution for enumerations from HH member respondents. The unweighted distributions are shown in Tables 9U and 10U in the Appendix.

For enumerations from proxy respondents with reported Hispanic ethnicity, the percentages of nonHispanic and Hispanic enumerations that were assigned PIKs are comparable at 62% and 61%, respectively. For enumerations from proxy respondents with imputed Hispanic ethnicity, the percentages of nonHispanic and Hispanic enumerations that the PVS assigned PIKs are 17% and 10%, respectively. When considering only those proxy enumerations that the PVS attempted to assign PIKs, the percentages are 47% for nonHispanic and 53% for Hispanic. The major difference arises from the whole person imputations that have all their characteristics imputed constitute 77% (1474.2/1909.1) of the nonHispanic and 85% (446.4/528.1) of the Hispanic values.

**Table 9.** Weighted distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **proxy** respondents by whether the enumeration contained reported race/Hispanic ethnicity information (shown in 1,000s)

<b>Reported</b>						
	<b>Not Hispanic</b>			<b>Hispanic</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	2,753.6	1,097.3	3,850.9	532.6	224.4	757.0
Erroneous residence	192.9	70.2	263.0	40.6	20.9	61.5
Unresolved residence	216.3	769.9	986.2	67.5	170.1	237.6
Insufficient info for CCM	132.1	56.0	188.1	30.9	8.5	39.4
<b>Subtotal</b>	3,294.9	1,993.4	5,288.2	671.6	423.9	1,095.5
	62%	38%	100%	61%	39%	100%
<b>PIK not attempted</b>						
Whole person imputation		0.0	0.0		0.0	0.0
<b>Total</b>	3,294.9	1,993.4	5,288.2	671.6	423.9	1,095.5
	62%	38%	100%	61%	39%	100%

<b>Imputed</b>						
	<b>Not Hispanic</b>			<b>Hispanic</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	298.4	251.6	550.0	41.3	36.1	77.3
Erroneous residence	27.1	18.7	45.8	5.9	4.7	10.6
Unresolved residence	44.9	148.5	193.4	8.7	36.4	45.1
Insufficient info for CCM	7.6	16.2	23.8	2.5	4.5	7.1
<b>Subtotal</b>	378.0	434.9	812.9	58.4	81.7	140.1
	47%	53%	100%	42%	58%	100%
<b>PIK not attempted</b>						
Whole person imputation		1,474.2	1,474.2		446.4	446.4
<b>Total</b>	378.0	1,909.1	2,287.2	58.4	528.1	586.5
	17%	83%	100%	10%	90%	100%

**Table 10.** Weighted distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **HH member** respondents by whether the enumeration contained reported race/Hispanic ethnicity information (shown in 1,000s)

<b>Reported</b>						
	<b>Not Hispanic</b>			<b>Hispanic</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	27,006.0	1,351.0	28,357.0	6,528.1	900.4	7,428.5
Erroneous residence	678.7	141.1	819.8	140.6	62.5	203.1
Unresolved residence	24.0	44.2	68.2	393.9	173.6	567.5
Insufficient info for CCM	29.6	3.5	33.1	204.5	23.6	228.1
<b>Subtotal</b>	27,738.3	1,539.7	29,278.1	7,267.2	1,160.1	8,427.3
	95%	5%	100%	86%	14%	100%
<b>PIK not attempted</b>						
Whole person imputation		0.0	0.0		0.0	0.0
<b>Total</b>	27,738.3	1,539.7	29,278.1	7,267.2	1,160.1	8,427.3
	95%	5%	100%	86%	14%	100%

<b>Imputed</b>						
	<b>Not Hispanic</b>			<b>Hispanic</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	515.1	116.1	631.2	272.9	30.7	303.5
Erroneous residence	16.5	8.2	24.7	8.1	3.2	11.3
Unresolved residence	24.0	44.2	68.2	16.3	8.0	24.3
Insufficient info for CCM	29.6	3.5	33.1	11.6	3.5	15.1
<b>Subtotal</b>	585.2	172.0	757.2	308.9	45.4	354.2
	77%	23%	100%	87%	13%	100%
<b>PIK not attempted</b>						
Whole person imputation		460.9	460.9		122.0	122.0
<b>Total</b>	585.2	632.9	1,218.1	308.9	167.4	476.2
	48%	52%	100%	65%	35%	100%

When we turn our attention to enumerations from HH member respondents, we see that the percentage of proxy enumerations with reported Hispanic ethnicity that received PIKs in the PVS assignment process is 95% for nonHispanics and 86% for Hispanics. For enumerations from HH member respondents with imputed Hispanic ethnicity values, the percentages of nonHispanic and Hispanic enumerations that the PVS assigned PIKs are 48% and 65%, respectively. When restricting the calculation to only those enumerations that the PVS attempted to assign PIKs, the percentage that received PIKs is 77% for nonHispanics and 87% for Hispanics. The whole person imputations appear to be the source of the difference since they are 73% (460.9/632.9) of the nonHispanic and 73% (122.0/167.4) of the Hispanic imputed values without PIKs.

Moving to the age variable, Table 11 contains the unweighted distributions of reported and imputed values for proxy and HH members by the categories: 0 to 17 years of age and age 18 or over. Table 11 shows that 33.5% of the enumerations from proxy respondents have a reported age while percentage of enumerations with reported values from HH member respondents is 86.0%.

**Table 11.** Weighted distributions of reported and imputed values of the age variable by whether the respondent was a proxy or a HH member

	reported	imputed	row total	column %
<b>Proxy</b>				
0 to 17 years	872.1	1,341.2	2,213.3	23.9%
row %	39.4%	60.6%	100.0%	
18 years or over	2,233.7	4,810.4	7,044.1	76.1%
row %	31.7%	68.3%	100.0%	
column total	3,105.8	6,151.6	9,257.4	100.0%
row %	33.5%	66.5%	100.0%	
<b>HH member</b>				
0 to 17 years	11,828.3	1,467.3	13,295.6	31.9%
row %	89.0%	11.0%	100.0%	
18 years or over	24,057.9	4,387.8	28,445.7	68.1%
row %	84.6%	15.4%	100.0%	
column total	35,886.2	5,855.1	41,741.2	100.0%
row %	86.0%	14.0%	100.0%	

Another consideration is whether the percentages in the categories 10 to 17 years and 18 years or over with reported values are comparable within proxy and HH member respondents. For the proxy responses, 39.4% of the values in the 0 to 17 category were reported while 31.7% of the 18 years or over values were reported. For HH member responses, 89.0% of the 0 to 17 values were reported, and 84.6% of the 18 years or over values were reported.

Table 12 shows the weighted distributions of the combined CCM residence status for proxy enumerations with reported and imputed values of age. Table 13 shows the same distributions for enumerations from HH member respondents. The Appendix contains the unweighted distributions in Tables 12U and 13U.

**Table 12.** Weighted distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **proxy** respondents by whether the enumeration contained age information

<b>Reported</b>						
	<b>0 to 17 years</b>			<b>18 years or over</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	711.6	29.3	740.9	1,757.2	79.3	1,836.5
Erroneous residence	37.6	4.4	42.0	136.2	13.1	149.3
Unresolved residence	35.1	11.0	46.1	131.9	22.7	154.6
Insufficient info for CCM	43.0	0.0	43.0	89.5	3.9	93.4
<b>Subtotal</b>	827.3	44.8	872.1	2,114.7	119.0	2,233.7
	95%	5%	100%	95%	5%	100%
<b>PIK not attempted</b>						
Whole person imputation		0.0	0.0	0	0.0	0.0
<b>Total</b>	827.3	44.8	872.1	2,114.7	119.0	2,233.7
	95%	5%	100%	95%	5%	100%

<b>Imputed</b>						
	<b>0 to 17 years</b>			<b>18 years or over</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	165.8	324.4	490.2	75.7	82.2	157.8
Erroneous residence	17.0	14.8	31.7	136.9	810.1	947.0
Unresolved residence	33.6	281.0	314.7	30.2	68.9	99.1
Insufficient info for CCM	10.4	12.3	22.7	0.0	1,438.7	1,438.7
<b>Subtotal</b>	226.8	632.5	859.3	242.8	2,399.9	2,642.7
	26%	74%	100%	9%	91%	100%
<b>PIK not attempted</b>						
Whole person imputation		481.9	481.9		1,438.7	1,438.7
<b>Total</b>	226.8	1,114.4	1,341.2	242.8	3,838.7	4,081.5
	17%	83%	100%	6%	94%	100%

**Table 13.** Weighted distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **HH member** respondents by whether the enumeration contained age information

<b>Reported</b>						
	<b>0 to 17 years</b>			<b>18 years or over</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	10,401.5	291.5	10,693.1	21,019.2	606.0	21,625.3
Erroneous residence	243.2	21.1	264.3	507.2	68.5	575.6
Unresolved residence	538.4	63.3	601.7	990.5	185.4	1,176.0
Insufficient info for CCM	264.2	5.0	269.2	661.1	19.9	681.0
<b>Subtotal</b>	11,447.4	380.9	11,828.3	23,178.1	879.8	24,057.9
	97%	3%	100%	96%	4%	100%
<b>PIK not attempted</b>						
Whole person imputation		0.0	0.0	0	0.0	0.0
<b>Total</b>	11,447.4	380.9	11,828.3	23,178.1	879.8	24,057.9
	97%	3%	100%	96%	4%	100%

<b>Imputed</b>						
	<b>0 to 17 years</b>			<b>18 years or over</b>		
<b>Census Day residence status</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>	<b>with PIK</b>	<b>without PIK</b>	<b>Total</b>
<b>PIK attempted</b>						
Correct residence	673.0	435.8	1,108.9	2,228.2	1,064.8	3,293.0
Erroneous residence	27.0	30.5	57.5	66.6	94.9	161.5
Unresolved residence	30.1	91.8	122.0	154.4	254.2	408.6
Insufficient info for CCM	12.4	12.4	24.8	53.2	42.7	95.9
<b>Subtotal</b>	742.5	570.6	1,313.1	2,502.4	1,456.6	3,959.0
	57%	43%	100%	63%	37%	100%
<b>PIK not attempted</b>						
Whole person imputation		154.2	154.2		428.7	428.7
<b>Total</b>	742.5	724.8	1,467.3	2,502.4	1,885.3	4,387.8
	51%	49%	100%	57%	43%	100%



For enumerations from proxy respondents with reported age, the percentage with a PIK in the 0 to 17 years category is 95%, which is equal to the percentage in the 18 years or over category. When age is imputed for a proxy enumeration, the percentages with PIKs of those that the PVS assigned PIKs are 17% for the 0 to 17 years category and 6% for the 18 years or over category. A contributor to the difference between the PIK rates for the enumerations with and without PIKs is the whole person imputations that have all their characteristics imputed and constitute 43% (481.9/1114.4) of the 0 to 17 years category and 37% (1,438.7/3,838.7) of the 18 years or over category.

Although the difference in the patterns of reporting Hispanic ethnicity and age by proxy respondents is very surprising, reported age may be a feature that makes a proxy enumeration more likely to be in the correct location. Table 14 shows the correct enumeration rates when Hispanic ethnicity and age are reported or imputed by type of NRFU respondents. For proxy respondents, the difference in the correct enumeration rate when age is reported versus imputed is 20% (83% - 63%). For reported versus imputed Hispanic ethnicity, the difference in correct enumeration rate for proxy responses is 6% (72% - 66%).

Evidently, proxy respondents reported Hispanic ethnicity at times when they did not report age. However, the proxy respondents who reported age may have been more accurate about who the Census Day residents than the proxies who did not report age. Possibly instructions to interviewers regarding proxy interviews did not place as much emphasis on collecting ages as was placed on other characteristics. A potential remedy is for interviewer training to emphasize obtaining birth dates from proxy respondents when possible and if the proxy respondent does not know birthdates, to obtain age. If the proxy respondents do not know birthdates or ages, the interviewers could ask the number of residents under 18 years of age and 18 years of age or over, which would provide more information about the number of children and adults. Then the design of the imputation procedure could preserve the number of children and adult groups.

**Table 14. Weighted correct enumeration (CE) rate for enumerations in occupied HUs in the combined CCM by whether Hispanic ethnicity and age are reported or imputed by type of NRFU respondent. (shown in 1,000's)**

Characteristic & reporting status	Proxy respondent			HH member respondent		
	Total	CE	% CE	Total	CE	%CE
<b>Hispanic ethnicity</b>						
Reported	6,384	4,608	72%	37,705	35,786	95%
Imputed	953	627	66%	1,111	935	84%
<b>Age</b>						
Reported	3,106	2,577	83%	35,886	32,318	90%
Imputed	4,231	2,658	63%	5,272	4,402	83%

## 6. Possibility of increasing number of ARs evaluated

Since the PIKs for 5,677 administrative records in NRFU households with proxy responses could not be found in the combined CCM records at the same address, we investigated whether the household-based matching used by CCM is a viable alternative matching procedure that possibly could find additional links. This analysis is unweighted because it is of an operational nature for this study.

Considering this option seems reasonable since there are E-sample and P-sample records that were not assigned a PIK, but the CCM's household-based matching operation was able to resolve their residence status. We do have to keep in mind that the CCM matching operation combined the electronic household-based matching algorithm with clerical operations. However, the CCM approach may have been able to link some records where the names and other information were somewhat different, so different that PIK matching could not make these matches. Similar differences in the names in the combined CCM and the Numident files may be the cause of the combined CCM records not receiving SSNs, and thereby, not being able to receive a PIK.

Although, as shown in Table 3, PIK-based matching was not able to link 3,447 of 5,677 AR records in proxy HUs to an enumeration anywhere in the census, the balance, 2,230 records, were found at other addresses. Table 15 shows the location where the matching PIK was found in the census relative to the address on the AR record. We see that 16.7% of the matched PIKs were in the same block, 11.2% were in the same tract, but not the same block, and 42.4% were in the same county but not the same tract. In addition, 15.9% were in the same state but not the same county while 13.8% were in a different state. . Table 15 also shows that the distribution of the enumerations matching HH member enumerations by distance from the AR address is comparable. We do not know how many of the remaining 3,447 ARs that PIK-based matching could not link would link to records in the combined CCM using household-based matching.

If one were to relax the requirement that the AR and combined CCM records link at the same address to allowing ARs to link to combined CCM records at different addresses within the same block, some additional residence codes can be assigned. Table 16 shows that of the ARs with links to different addresses in the same block, 71.0% of the ARs in proxy HUs and 78.9% of those in HH member HUs link to combined CCM records. Subsampling in large blocks caused CCM not to process some enumerations causing some not to receive CCM residence status codes. The relaxed linking shows that 61.1% of the ARs in proxy HUs and 71.4% of the ARs in HH member HUs that link at a different address would be considered correct.

**Table 15.** Location of census enumeration that matched to AR record in a NRFU HU that has a PIK not found in CCM records at the same address by type of NRFU respondent

Status	Distance from Other Unit					Total	%
	Same Block	Same Tract	Same County	Same State	Outside State		
<b>Proxy respondent</b>							
<u>PIK found at another address</u>							
Number of addresses							
1	340	233	903	342	290	2,108	37.1%
2+	33	17	43	12	17	122	2.1%
subtotal	373	250	946	354	307	2,230	39.3%
% at location	16.7%	11.2%	42.4%	15.9%	13.8%	100.0%	
<u>PIK not found at another address</u>						3,447	60.7%
Total						5,677	100.0%
<b>HH member respondent</b>							
<u>PIK found at another address</u>							
Number of addresses							
1	879	426	2,161	963	651	5,080	42.8%
2+	68	19	106	31	21	245	2.0%
subtotal	947	445	2,267	994	672	5,325	44.8%
% at location	17.8%	8.4%	42.6%	18.7%	12.6%	100.0%	
<u>PIK not found at another address</u>						6,564	55.2%
Total						11,882	100.0%

Next, we need to consider how many combined CCM records that were not linked to ARs at the same address or with their block, and therefore, might link using household-based matching. Table 17 shows that there are 5,776 resolved combined CCM records in NRFU HUs with proxy responses do not link to ARs through PIK-based matching. Of these, 2,331 do not have PIKs and 3,445 have PIKs. Approximately 1,000 of the records in the P-sample that are not in the census at the same address in NRFU HUs with proxy respondents are in-movers who were not at the address on Census Day. The corresponding number of in-movers for NRFU HUs with HH member respondents is approximately 1,500.

**Table 16.** Residence status of combined CCM record that links to an AR record in the same block but with a different address

<b>Census Day residence status</b>	<b>Proxy</b>		<b>HH member</b>	
<b>Linked to different address in CCM sample</b>				
Correct residence (block)	228	61.1%	676	71.4%
Erroneous residence (block)	16	4.3%	37	3.9%
Unresolved residence (block)	21	5.6%	34	3.6%
<b>Subtotal for Linked</b>	265	71.0%	747	78.9%
<b>Linked at different address not in CCM sample*</b>	108	29.0%	200	21.1%
<b>total</b>	373	100.0%	947	100.0%

\*Subsampling of large blocks caused some enumerations not to received CCM codes.

**Table 17.** Persons in E-sample and/or P-sample with a resolved CCM status that were not found in ARs at the same address using matching with PIKs by whether they were assigned a PIK and by respondent

<b>Source</b>	<b>Proxy</b>	<b>HH member</b>
<b>In CCM, no PIK</b>		
In census	2,331	3,630
In P-sample	1,646	5,105
<b>In CCM with PIK, but not found in ARs at same address</b>		
In census & in P-sample	504	5,485
In census but not in P-sample	598	4,158
In P-sample but not in CUF	2,343	4,388
subtotal	3,445	14,031
<b>Total resolved CCM, but not linked to AR at the same address</b>	<b>5,776</b>	<b>17,661</b>

On the other hand, an advantage of using the CCM is there were two opportunities to list the names and characteristics for the Census Day residents of a HU, namely the census and the CCM personal interview. If one interview does not obtain sufficient information to identify the residents, then the other interview may be more successful. As evidence of this occurring, Table 18 shows location of the matching census enumeration for P-sample matched nonmovers who were assigned PIKs but the PIKs were not found in the census. Interestingly, the percentage found at the same address, denoted by MAFID, is essentially equal at 89.5% for proxy respondents and 89.2% for HH member respondents. The difference appears in the percentages that were matched to enumerations with blank names, which is 34.9% for proxy respondents and 10.3% for HH member respondents. These matches were identified by clerical matching

rather than computer matching. Those linking to enumerations with a name where the HU is not in the E-sample appear to arise from a difference in the subsampling for the P and E samples in a few large block clusters caused some HUs to be included in the P sample but not the E sample or vice versa.

**Table 18.** Location of matching census enumerations found by CCM for P-sample matched nonmovers with PIKs, but the PIKs were not found in the census at the same address by type of NRFU respondent

Category		Proxy		HH Member	
Same MAFID	Blank Name	424	34.9%	182	10.3%
	Filled Name	664	54.6%	1,391	78.9%
	No PIK	561		671	
	Same PIK, not E-sample*	93		701	
	Match to duplicate	10		19	
Total		1,088	89.5%	1,573	89.2%
Same Block		83	6.8%	131	7.4%
Same Tract		11	0.9%	18	1.0%
Same County		27	2.2%	21	1.2%
Same State		4	0.3%	16	0.9%
Outside State		3	0.2%	3	0.2%
Total		1,216	100.0%	1,762	100.0%

\*A difference in the subsampling for the P and E samples in a few large block clusters caused some HUs to be included in the P sample but not the E sample or vice versa.

There definitely is a possibility of household-based matching being able to link additional ARs to combined CCM records with resolved residence status. The number of additional links that might be obtained is hard to estimate with the information we have at this point. The cost of undertaking such a project has to be weighed against the potential for additional information.

## 7. Summary and Next Steps

To conclude, we return to our research questions.

### **1) Are proxy responses more or less accurate than administrative records?**

Answering this question is not as straightforward as it sounds. Our investigation used the 10,416 NRFU HUs with proxy respondents and 16,876 NRFU HUs with HH member respondents in both the CCM P-sample and E-sample that the census classified as occupied. We studied the AR and census records assigned to the addresses for these HUs and used the combined CCM records to evaluate the accuracy of the records from each source.

The major findings from our study follow:

- Approximately half of the NRFU proxy HUs do not have ARs in the IRS 1040 and Medicare files for all of 2010. These two administrative sources include some information on household composition. Unless additional high-quality AR sources that cover these addresses can be found, these HUs without ARs will need contact by NRFU enumerators or whole HH imputation.
  - Increasing the number of AR sources elevates the importance of identifying duplicate records across HUs and developing rules for which address to keep as the person's Census Day address. A related issue is processing the files in manner that identifies and removes people who died before Census Day or were born after Census Day. CCM finds a few such census enumerations and codes them as erroneous.
- By almost any standard, proxy enumerations that can be assigned PIKs tend to be in the correct location. Therefore, one indicator for a higher quality NRFU enumeration appears to be whether it has enough information for the Census Bureau's PVS algorithm to assign a PIK.
  - Many data-defined census enumerations that meet the CCM criteria of sufficient information, which is a name and two characteristics, could not be assigned PIKs but were found by CCM to be at the correct location.
  - Whole household imputations are 20.7% of enumerations in proxy HUs and 1.4% of enumerations in HH member HUs.
- When the NRFU enumerations had enough information for the PVS system to attempt to assign a PIK, the percentages that received PIKs were 73% of the proxy enumerations and 92% of the HH member enumerations.
  - When the whole person imputations are included in the denominator, the percentages receiving PIKs drop to 58% for proxy enumerations and 91% for HH member.
  - Possibly NRFU enumerations could be PIK-ed as they come in during NRFU for a quality assessment.
- The combined CCM found that an unweighted 32% of proxy HUs with ARs had perfect HH composition. That means that all the AR persons in the HU were Census Day residents and no Census Day residents were omitted from the AR roster. The enumerations with unresolved residence status were not considered to be at the correct location although some likely are but there is not enough information to make a determination.
  - Household-based matching may be able to increase the number of links to combined CCM records over what was linked using PIK-based matching.
- When focusing only on population count, the percentage of HUs have an AR count that agrees with the census count is an unweighted 51% among HUs with proxy respondents and among HUs with HH member respondents.
  - Unweighted, 34% of proxy HUs have an AR count that agrees with the census count and all the AR PIKs in combined CCM and while 44% of HH member HUs meet the same criteria.
- Duplication may be a problem when using ARs to enumerate whole HHs. Our study found that an unweighted 17% of ARs in proxy HUs and 11% of ARs in HH member HUs linked to a census enumeration at an address other than the AR address. Also troubling is that an unweighted 27% of ARs in proxy HUs and 12% of ARs in HH member HUs did not link anywhere in the census.
  - For ARs that link to a census enumeration and address other than the AR address,
    - Using the AR at its address may create a duplicate in the census. Census operations may need to search census enumerations, particularly self-responses,

to be sure that an AR enumeration is not a duplicate. The addition of questions regarding other residences to the census questionnaire may aid in avoiding duplicates.

- The use of household-based matching between the ARs that link to a census address other than their AR addresses and the combined CCM has the potential for finding more links. If there is a link, an examination may provide information about the reason for the person being at both addresses. If no link, then the enumeration outside sample block could be person's only enumeration.
- An examination of item nonresponse for Hispanic ethnicity found that 69.0% of proxy enumerations and 95.9% of HH member enumerations have reported Hispanic ethnicity values, relatively high since the number of imputed values includes whole person imputations.
- An examination of item nonresponse for age found that 33.5% of proxy enumerations and 86.0% of HH member enumerations have reported age, surprisingly low. This is a concern because birthdate and/or age are important in enabling the PVS to assign a PIK to a record. However, reported age may be a feature that makes a proxy enumeration more likely to be in the correct location since the correct enumeration rate was 83% among enumerations with reported age versus 63% for enumerations with imputed age.
  - Possibly proxy respondents are more willing to report Hispanic ethnicity than age.
  - NRFU enumerators could ask for number of children & adults rather than just population count.

The recommendations arising from our findings are as follows:

- The design of NRFU operations would profit by including consideration of strategies to obtain high-quality proxy responses, possibly secondary in priority to strategies to obtain responses from HH members. Such strategies include developing contact tactics that incorporate times when knowledgeable proxy respondents are likely to be accessible, namely at home for neighbors or on the premises for multi-unit building managers. In addition, design the training of interviewers to emphasize that the name and age of the residents from proxy respondents are priorities.
  - NRFU proxy respondents who can provide a name and two characteristics appear to report high quality information and therefore, should be considered better than census whole person imputations.
  - Among the enumerations that were not whole person imputations, 60% of the enumerations in proxy HUs 91% for enumerations in HH member HUs could be assigned PIKs.
  - Since the percentage of the CCM HUs with proxy respondents that did have ARs in IRS 1040 and Medicare files for all of 2010 is 51%, ARs cannot be considered a cure-all at this point in time. Unless additional high-quality sources of ARs can be found for the 49% with no ARs in these IRS and Medicare files, NRFU interviews or whole person imputations are the only other alternatives for these HUs. However, using ARs strategically can save money during NRFU.
- Developing methodology for identifying the HUs where the ARs are perfect has the potential for several uses in NRFU that would reduce cost and improve quality. For example, the ARs could be

used to enumerate the HU and by reducing the number of visits by enumerators to the HU save resources. When sex and age of the residents is missing from the enumeration, these could be pulled from ARs rather than imputed.

- CCM found that the administrative records for an unweighted 33% of proxy HUs with ARs had a perfect enumeration, meaning all the ARs represented Census Day residents at the address and no Census Day residents were omitted from the AR list for the HU.
- Using ARs for enumeration purposes requires development of NRFU procedures for avoiding the creation of duplicate enumerations. Since NRFU enumerations also have the potential to be duplicate enumerations, census operations need to handle NRFU and AR enumerations in a consistent manner. Consistent treatment of NRFU and AR enumerations of people who died before Census Day or who were born after Census Day is also important.
  - An unweighted 17% of the proxy enumerations and 11% of the HH member enumerations link to census enumerations at other addresses, most of which are outside of the sample block cluster, which raises a concern that using these ARs for enumeration purposes would create duplicate enumerations.
  - The IRS 1040 and Medicare files used in this study contained 88 records in proxy HUs and 237 records in HH member HUs for people who died in 2009.
- Household-based matching between the ARs with no links and the combined CCM records may identify more links and provide more insight into the quality of the ARs and reasons for the failure to link.
  - An unweighted 27% of the ARs in proxy HUs and 12% of the ARs in HH member HUs do not link to a combined CCM record or to a census record anywhere in the census, which means their quality cannot be evaluated.
  - Although 21% of the ARs in proxy HUs are whole person imputations, only 2% of the enumerations in HH member HUs are whole person imputations, implying that some, but definitely not all, of the failures to link are explained by whole person imputations.

## **2)How does the quality of proxy responses vary?**

The next steps in our research will concentrate on investigating how the quality of the proxy responses may vary. In further investigations, we will examine the characteristics of the HUs where the combined CCM found their individual ARs to be perfect, that is, the exact household members were correctly enumerated versus those HUs with ARs that had errors or could not be evaluated.

We also will investigate the relationship between number of prior contact attempts and correct proxy responses and identify characteristics of HUs with complete correct admin records among NRFU proxy responses. By characteristics, we mean operational as well as geographic, socio-economic variables. We plan to merge some ACS data for this investigation. The analytical methods we plan to use include decision trees and other multivariate statistical methodologies.

Another application for the methodology in this paper is the evaluation of the results of the predictive modeling approach used in the 2015 Census Test to identify HUs for AR enumeration (Morris 2015).



Such a study would apply the predictive modeling approach to the NRFU HUs in the combined CCM with ARs to produce scores indicating the ones that are most likely to be correct enumerations. Then the results of the combined CCM would provide a measure of accuracy such as the ones used in Table 7, namely AR Perfect, AR Omissions, and AR E&U. In addition, the results would provide guidance for identifying cut-offs for the scores from the predictive modeling that yield the highest percentages of HUs whose ARs fall in the category of AR Perfect.

## 6. References

- Keller, A. and T. Fox (2012) “2010 Census Coverage Measurement Estimation Report: Components of Census Coverage for the Household Population in the United States.” DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES #2010-G-04. U.S. Census Bureau. Washington, DC.
- Lohr, S. (1999) *Sampling: Design and Analysis*. Cengage Learning. Boston, MA.
- Mule, T. (2012) “2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States.” DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES #2010-G-01. U.S. Census Bureau. Washington, DC.
- Morris, D. S. (2015) “An Approach to Using Administrative Records in the 2020 Census.” In *2015 Joint Statistical Meeting Proceedings*. American Statistical Association. Alexandria, VA:
- Mulry, M. and Keller, A. (2015) “Are Proxy Responses Better Than Administrative Records?” In *2015 Joint Statistical Meeting Proceedings*. American Statistical Association. Alexandria, VA:
- Mulry, M. H. and Spencer, B. D. (2012) “A Framework for Cost Models Relating Cost and Data Quality.” Invited presentation at the 2012 International Total Survey Error Workshop. Sanpoort, The Netherlands, September 2 - 4 2012.  
[http://www.niss.org/sites/default/files/Mulry\\_september2012.pdf](http://www.niss.org/sites/default/files/Mulry_september2012.pdf)
- Olson, D. and Griffin, R. (2012) “2010 Census Coverage Measurement Estimation Report: Aspects of Modeling.” DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES #2010-G-10. U.S. Census Bureau. Washington, DC.
- Wagner, D., & Layne, M. (2014). The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications’ (CARRA) Record Linkage Software CARRA Working Paper Series. #2014-01 Last accessed 9/22/2014 at [https://www.census.gov/srd/carra/CARRA\\_PVS\\_Record\\_Linkage.pdf](https://www.census.gov/srd/carra/CARRA_PVS_Record_Linkage.pdf)

## Appendix

**Table 3U. Unweighted** distributions of combined CCM residence status for enumerations and ARs in NRFU HUs in the combined CCM by NRFU respondent type

Census Day residence status	Proxy respondent			
	NRFU		AR	
	count	%	count	%
<b>Correct residence</b>	6,637	56.4%	6,191	48.1%
<b>Erroneous residence</b>	481	4.1%	519	4.0%
<b>Unresolved residence</b>	1,850	15.7%	493	3.8%
<b>NRFU not processed by CCM</b>				
<b>Insufficient info</b>	290	2.5%	-	-
<b>Whole person Imputation</b>	2,508	21.3%	-	-
<b>AR PIK not in census at same address</b>				
<b>Found at another census address</b>	-	-	2,230	17.3%
<b>Not linked to census records</b>	-	-	3,447	26.8%
	11,766	100.0%	12,880	100.0%

Census Day residence status	HH member respondent			
	NRFU		AR	
	count	%	count	%
<b>Correct residence</b>	45,018	87.4%	36,084	70.9%
<b>Erroneous residence</b>	1,392	2.7%	1,258	2.5%
<b>Unresolved residence</b>	3,042	5.9%	1,645	3.2%
<b>NRFU not processed by CCM</b>				
<b>Insufficient info</b>	1,285	2.5%	-	-
<b>Whole person Imputation</b>	748	1.5%	-	-
<b>AR PIK not in census at same address</b>				
<b>Found at another census address</b>	-	-	5,318	10.5%
<b>Not linked to census records</b>	-	-	6,564	12.9%
	51,485	100.0%	50,869	100.0%

**Table 4U. Unweighted** distributions of combined CCM residence status for enumerations in NRFU HUs by NRFU respondent type and PIK status.

Census Day residence status	NRFU Proxy			NRFU HH member		
	with PIK	without PIK	Total	with PIK	without PIK	Total
<b>PIK attempted</b>						
Correct residence	4,585	2,052	6,637	41,832	3,186	45,018
Erroneous residence	317	164	481	1,088	303	1,391
Unresolved residence	449	1,401	1,850	2,209	834	3,043
Insufficient info for CCM	190	100	290	1,178	107	1,285
<b>Subtotal</b>	5,541	3,717	9,258	46,307	4,430	50,737
	60%	40%	100%	91%	9%	100%
<b>PIK not attempted</b>						
Whole person imputation		2,508	2,508		748	748
<b>Total</b>	5,541	6,225	11,766	46,307	5,178	51,485
	47%	53%	100%	90%	10%	100%

**Table 8U.** Distributions of reported and imputed values of the Hispanic variable by whether the respondent was a proxy or a HH member (unweighted)

	reported	imputed	row total	column %
<b>Proxy</b>				
Hispanic	1,489	782	2,271	19.3%
row %	65.6%	34.4%	100.0%	
not Hispanic	6,548	2,947	9,495	80.7%
row %	69.0%	31.0%	100.0%	
column total	8,037	3,729	11,766	100.0%
row %	68.3%	31.7%	100.0%	
<b>HH member</b>				
Hispanic	10,963	1,497	12,460	24.2%
row %	88.0%	12.0%	100.0%	
not Hispanic	38,376	649	39,025	75.8%
row %	98.3%	1.7%	100.0%	
column total	49,339	2,146	51,485	100.0%
row %	95.8%	4.2%	100.00%	

**Table 9U.** Distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **proxy** respondents by whether the enumeration contained reported race/Hispanic ethnicity information (**unweighted**)

Combined CCM status	Hispanic ethnicity not imputed						Hispanic ethnicity imputed					
	not Hispanic		Hispanic		combined		not Hispanic		Hispanic		combined	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
<b>Census Day residence correct</b>	4,777	73.0	1,061	71.3	5,838	72.6	685	23.2	114	14.6	799	21.4
with PIK	3,408	52.0	732	49.2	4,140	51.5	383	13.0	62	7.9	445	11.9
without PIK	1,369	20.9	329	22.1	1,698	21.1	302	10.2	52	6.6	354	9.5
<b>Census Day residence erroneous</b>	331	5.1	72	4.8	403	5.0	64	2.2	14	1.8	78	2.1
with PIK	230	3.5	47	3.2	277	3.4	31	1.1	9	1.2	40	1.1
without PIK	101	1.5	25	1.7	126	1.6	33	1.1	5	0.6	38	1.0
<b>Census Day residence unresolved</b>	1,440		356		1,796		277		67		344	
with PIK	430	6.6	121	8.1	551	6.9	72	2.4	16	2.0	88	2.4
sufficient info	290	4.4	84	5.6	374	4.7	64	2.2	11	1.4	75	2.0
insufficient info	140	2.1	37	2.5	177	2.2	8	0.3	5	0.6	13	0.3
without PIK	1,010	15.4	235	15.8	1,245	15.5	205	7.0	51	6.5	256	6.9
sufficient info	947	14.5	225	15.1	1,172	14.6	185	6.3	44	5.6	229	6.1
insufficient info	63	1.0	10	0.7	73	0.9	20	0.7	7	0.9	27	0.7
<b>Whole person imputation</b>	0	0	0		0	0	1,921	65.2	587	75.1	2,508	67.3
<b>Total</b>	6,548	100.0	1,489	100.0	8,037	100.0	2,947	100.0	782	100.0	3,729	100.0

**Table 10U.** Distribution of CCM residence status for enumerations in NRFU HUs for **HH member** respondents by whether the enumeration contained race/Hispanic ethnicity information (**unweighted**)

Combined CCM status	Hispanic ethnicity not imputed						Hispanic ethnicity imputed																	
	not Hispanic		Hispanic		combined		not Hispanic		Hispanic		combined													
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%												
<b>Census Day residence correct</b>		34,251		89.3		9,599		87.6		43,850		88.9		760		50.8		408		62.9		1,168		54.4
with PIK	32,461		84.6		8,402		76.6		40,863		82.8		607		40.5		362		55.8		969		45.2	
without PIK	1,790		4.7		1,197		10.9		2,987		6.1		153		10.2		46		7.1		199		9.3	
<b>Census Day residence erroneous</b>		1,051		2.7		290		2.6		1,341		2.7		36		2.4		14		2.2		50		2.3
with PIK	856		2.2		197		1.8		1,053		2.1		25		1.7		10		1.5		35		1.6	
without PIK	195		0.5		93		0.8		288		0.6		11		0.7		4		0.6		15		0.7	
<b>Census Day residence unresolved</b>		3,074				1,074				4,148				123				57				180		
with PIK		2,496		6.5		789		7.2		3,285		6.7		68		4.5		34		5.2		102		4.8
sufficient info	1,606		4.2		544		5.0		2,150		4.4		37		2.5		22		3.4		59		2.7	
insufficient info	890		2.3		245		2.2		1,135		2.3		31		2.1		12		1.8		43		2.0	
without PIK		578		1.5		285		2.6		863		1.7		55		3.7		23		3.5		78		3.6
sufficient info	516		1.3		249		2.3		765		1.6		52		3.5		17		2.6		69		3.2	
insufficient info	62		0.2		36		0.3		98		0.2		3		0.2		6		0.9		9		0.4	
<b>Whole person imputation</b>		0		0		0		0		0		0		578		38.6		170		26.2		748		34.9
<b>Total</b>		38,376		100.0		10,963		100.0		49,339		100.0		1,497		100.0		649		100.0		2,146		100.0

**Table 11U.** Distributions of reported and imputed values of the age variable by whether the respondent was a proxy or a HH member (unweighted)

	<b>reported</b>	<b>imputed</b>	<b>Row total</b>	<b>column %</b>
<b>Proxy</b>				
0 to 17 years	1,091	1,706	2,797	23.8%
row %	39.0%	61.0%	100.0%	
18 years or over	2,793	6,176	8,969	76.2%
row %	31.1%	68.9%	100.0%	
column total	3,884	7,882	11,766	100.0%
row %	33.0%	67.0%	100.0%	
<b>HH member</b>				
0 to 17 years	14,394	1,890	16,284	31.6%
row %	88.4%	11.6%	100.0%	
18 years or over	29,539	5,662	35,201	68.4%
row %	83.9%	16.1%	100.0%	
column total	43,933	7,552	51,485	100.0%
row %	85.3%	14.7%	100.00%	

**Table 12U.** Distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **proxy** respondents by whether the enumeration contained age information (**unweighted**)

2010 proxy responses unweighted

Combined CCM status	Age not imputed						Age imputed																	
	0 to 17 years of age		18 years of age or over		all ages		0 to 17 years of age		18 years of age or over		all ages													
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%												
<b>Census Day residence correct</b>		921		84.4		2,303		82.5		3,224		83.0		642		37.6		2,771		44.9		3,413		43.3
<b>with PIK</b>	889		81.5		2,200		78.8		3,089		79.5		215		12.6		1,281		20.7		1,496		19.0	
<b>without PIK</b>	32		2.9		103		3.7		135		3.5		427		25.0		1,490		24.1		1,917		24.3	
<b>Census Day residence erroneous</b>		58		5.3		177		6.3		235		6.1		39		2.3		207		3.4		246		3.1
<b>with PIK</b>	52		4.8		159		5.7		211		5.4		19		1.1		87		1.4		106		1.3	
<b>without PIK</b>	6		0.5		18		0.6		24		0.6		20		1.2		120		1.9		140		1.8	
<b>Census Day residence unresolved</b>		112				313				425				402				1,313				1,715		
<b>with PIK</b>		97		8.9		281		10.1		378		9.7		50		2.9		211		3.4		261		3.3
<b>sufficient info</b>	52		4.8		183		6.6		235		6.1		38		2.2		176		2.8		214		2.7	
<b>insufficient info</b>	45		4.1		98		3.5		143		3.7		12		0.7		35		0.6		47		0.6	
<b>without PIK</b>		15		1.4		32		1.1		47		1.2		352		20.6		1,102		17.8		1,454		18.4
<b>sufficient info</b>	15		1.4		29		1.0		44		1.1		336		19.7		1,021		16.5		1,357		17.2	
<b>insufficient info</b>	0		0.0		3		0.1		3		0.1		16		0.9		81		1.3		97		1.2	
<b>Whole person imputation</b>		0		0		0		0		0		0		623		36.5		1,885		30.5		2,508		31.8
<b>Total</b>		1,091		100.0		2,793		100.0		3,884		100.0		1,706		100.0		6,176		100.0		7,882		100.0

**Table 13U.** Distribution of CCM residence status for enumerations in NRFU HUs in the combined CCM for **HH member** respondents by whether the enumeration contained age information (**unweighted**)

2010 HH member unweighted

Combined CCM status	Age not imputed						Age imputed					
	0 to 17 years of age		18 years of age or over		all ages		0 to 17 years of age		18 years of age or over		all ages	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
<b>Census Day residence correct</b>	12,977	90.2	26,421	89.4	39,398	89.7	1,384	73.2	4,236	74.8	5,620	74.4
with PIK	12,592	87.5	25,598	86.7	38,190	86.9	834	44.1	2,808	49.6	3,642	48.2
without PIK	385	2.7	823	2.8	1,208	2.7	550	29.1	1,428	25.2	1,978	26.2
<b>Census Day residence erroneous</b>	350	2.4	748	2.5	1,098	2.5	77	4.1	215	3.8	292	3.9
with PIK	314	2.2	658	2.2	972	2.2	32	1.7	84	1.5	116	1.5
without PIK	36	0.3	90	0.3	126	0.3	45	2.4	131	2.3	176	2.3
<b>Census Day residence unresolved</b>	1,067		2,370		3,437		215		677		892	
with PIK	977	6.8	2,083	7.1	3,060	7.0	59	3.1	268	4.7	327	4.3
sufficient info	670	4.7	1,294	4.4	1,964	4.5	42	2.2	203	3.6	245	3.2
insufficient info	307	2.1	789	2.7	1,096	2.5	17	0.9	65	1.1	82	1.1
without PIK	90	0.6	287	1.0	377	0.9	156	8.3	409	7.2	565	7.5
sufficient info	83	0.6	261	0.9	344	0.8	139	7.4	352	6.2	491	6.5
insufficient info	7	0.0	26	0.1	33	0.1	17	0.9	57	1.0	74	1.0
<b>Whole person imputation</b>	0	0	0		0	0	214	11.3	534	9.4	748	9.9
<b>Total</b>	14,394	100.0	29,539	100.0	43,933	100.0	1,890	100.0	5,662	100.0	7,552	100.0