Using a Child-to-Parent Linking Dataset in Administrative Records Modeling for Census Nonresponse Followup

Scott Konicki
U.S. Census Bureau[1]

Abstract
Preparing for the 2020 Census, the Census Bureau is researching the use of administrative records information for enumeration to reduce the number of field visits during the Nonresponse Followup operation. One of the concerns stemming from this research is the possible undercoverage of children in administrative data sources. In recent mid-decade tests, predictive models have been used to identify units with administrative records information of sufficient quality for enumeration. In this paper, we extend the predictive modeling approach by incorporating data from a child-to-parent linking dataset. This unique dataset associates children with their mother and father using data from the Social Security Administration. We explore methods of using these associations to include children on the administrative records roster for households where the children do not otherwise appear in the administrative records sources. We evaluate the impact of these alternative methods by assessing quality metrics using the 2010 Census data.

Key Words: Administrative Records, Nonresponse Followup, Decennial Census, Statistical Modeling

1. Introduction

The goal of the 2020 Census is to count each person only once and in the correct location. This is to be achieved at a lower cost per household (adjusted for inflation) than the 2010 Census while maintaining the same data quality. To meet this goal, the Census Bureau has researched fundamental changes to the design, implementation, and management of the 2020 Census. One of the major innovation research areas for the 2020 Census has been approaches to incorporate administrative records (AR) and third-party data into the census design. The AR Modeling Team has been researching the use of AR in the Nonresponse Followup (NRFU) operation. We have developed predictive models to identify addresses with high-quality AR for determining an occupied, vacant, or non-existent status. This process is described in detail in Morris et al. (2016) and Mule et al. (2016).

In this paper, we address the potential underrepresentation of children in AR sources by incorporating the Center for Administrative Records Research and Applications[2] (CARRA) Kidlink file as a core source for AR enumeration. The CARRA Kidlink file is a research file that attempts to link children to their parents using data from Social Security number applications. The CARRA Kidlink file offers an opportunity to improve the coverage of the core AR sources by associating children with their parents on the AR rosters.

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

Section 2 provides background information on the core AR sources and the Center for Administrative Records Research and Applications (CARRA) Kidlink file. Section 3 describes the methodology for incorporating the CARRA Kidlink file in the AR roster building and modeling processes. Section 4 presents results of this research with comparisons to the Baseline approach that does not use the CARRA Kidlink file. The results are based on a retrospective analysis using the 2010 Census.

2. Background

2.1. Core AR Sources

Part of the modeling process is the construction of household rosters using AR sources. Our research has used four core AR sources to construct the household rosters:

1. Internal Revenue Service (IRS) 1040 tax filings
2. IRS 1099 information returns
3. Center for Medicare and Medicaid Services (CMS) Medicare Enrollment Database (MEDB)
4. Indian Health Service (IHS) Patient Database

For units that are determined to be occupied from administrative records (henceforth referred to as AR occupied units), the AR persons from these sources are enumerated at the address. While children can appear on any of these sources, the vast majority of children are found only in the IRS 1040 source. IRS 1099 information returns are person-level records such as bank account interest statements and W-2 wage statements. These records typically apply to adults. Therefore, if we identify a unit as AR occupied based on the adult(s) from IRS 1099 returns and there is not an IRS 1040 filing for the address, then there is the possibility that our AR roster is missing other persons in the unit, especially children.

2.2. CARRA Kidlink Research File

The purpose of the CARRA Kidlink research file was to create a child-to-parent linking dataset using data from Social Security Number (SSN) applications. For each child on the Social Security Administration Numerical Identification (Numident) file, the CARRA attempted to assign a mother Protected Identification Key (PIK) and father PIK using the names of the parents from the SSN applications. A PIK is an encryption of a Social Security Number or Tax Identification Number (TIN) and is used as a unique identifier for matching person records between various AR and census sources. Note that for the child SSN applications used to build the CARRA Kidlink file, SSNs were not available for the parent records. The CARRA attempted to assign a PIK for the parents by matching the parent names to a set of reference files. Because this matching was not always successful, not all parent records could be assigned a PIK. There may also have been errors in the parent PIK assignment. For general information on how the Census Bureau assigns and uses PIKs, see Wagner and Layne (2014). For a discussion of the specific methodology for creating a previous iteration of the CARRA Kidlink research file, see Luque and Wagner (2015).

The CARRA Kidlink file used in this analysis contained 85 million records for births between 1996 and 2015, just over 4 million children per year. Table 1 shows the success rate of assigning a PIK to the mother and the father of the child. For about two-thirds of the children, a PIK could be assigned to both the mother and the father. Another 21

percent of children had only the mother PIK assigned. It may be that there was no father information available for these records.

Table 1: Parent PIK Assignment Rate

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

2.3. Research Approaches

Figure 1 shows the Nonresponse Followup (NRFU) contact strategy involving AR determinations that was used for the 2016 Census Test. This process is described in detail in Morris et al. (2016) and Mule et al. (2016). In this document, we discuss three potential uses of the CARRA Kidlink data during NRFU. The figure displays the point in the process for each approach.

1. *Using CARRA Kidlink to build AR rosters from the beginning*. Treat CARRA Kidlink as another core AR source and use the file to build AR rosters from the beginning. This approach uses CARRA Kidlink information in the green box to determine the AR occupied units and build the household rosters.
2. *Adding CARRA Kidlink children to rosters after the AR occupied identification.* Conduct the AR modeling using only the four core AR sources (Baseline approach). Then, add children from the CARRA Kidlink file to the roster for AR occupied units. This approach uses CARRA Kidlink information in the purple box to add children to units that are resolved by the AR occupied determination.
3. *Identifying units to receive the full NRFU contact strategy*. Use CARRA Kidlink to identify AR occupied units that may be missing children. Instead of adding children to the household roster, we would instead send units that have additional children from CARRA Kidlink to the field for the full contact strategy. These units would not be treated as AR occupied. This approach uses CARRA Kidlink information in the green box to determine the set of AR occupied units.

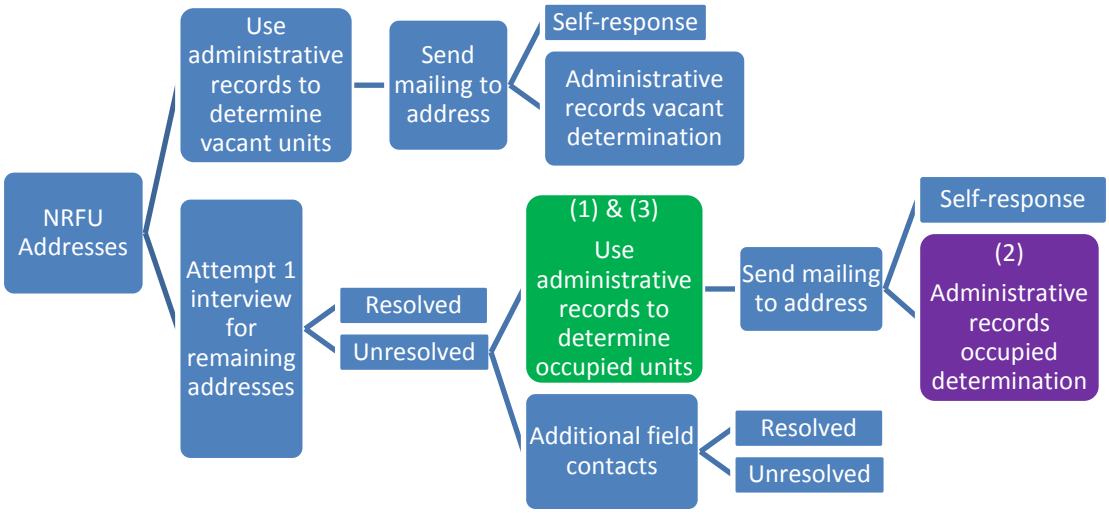

Figure 1: NRFU Contact Strategy for the 2016 Census Test with Potential Uses of CARRA Kidlink File

3. Methodology

Unlike the four core AR sources, the CARRA Kidlink file does not provide address information; the file only provides the child PIK and the PIKs of the parents, if they could be assigned. To place children from the CARRA Kidlink file at an address, we added the child to the AR roster for each of the addresses in which either the mother or the father appeared. If the mother and father appear in AR at separate addresses, then we added the child to each of these addresses. If a given parent appeared in multiple addresses in the core AR sources, then we added the child to each of those addresses.

The following example shows how the CARRA Kidlink file was used to build the household rosters. Here is the original AR person file built using the four core AR sources. An "X" indicates that a person was found at the address on a given AR source. Note that Mary Jones is found at two addresses in AR.

Address	Person	On IRS 1040	On IRS 1099	On MEDB	On IHS
101 Elm	Jack Smith		X		
101 Elm	Jill Smith		X		
202 Main	Mary Jones	X			
202 Main	Mike Hernandez			X	
303 Oak	Sherri Walker		X		

Next, suppose we observe the following two children on the CARRA Kidlink file. For Johnny Smith, both parents were assigned a PIK. For Annie Jones, only the mother could be assigned a PIK.

Child	Mother	Father
Johnny Smith	Jill Smith	Jack Smith
Annie Jones	Mary Jones	N/A

Johnny's parents are found at 101 Elm, so we add Johnny to the address as being "on Kidlink." Since Johnny's parents are not found at another AR address, he is not "on Kidlink elsewhere." We find Annie's mother Mary at 202 Main, so we add Annie to the AR roster for that address. Mary is also found at 303 Oak, so we add Annie to that address as well. Thus, Annie is "on Kidlink" and "on Kidlink elsewhere" for both units.

Address	Person	On IRS 1040	On IRS 1099	On MEDB	On Kidlink	On Kidlink elsewhere
101 Elm	Jack Smith		X			
101 Elm	Jill Smith		X			
101 Elm	Johnny Smith				X	
202 Main	Mary Jones	X				
202 Main	Mike Hernandez			X		
202 Main	Annie Jones				X	X
303 Oak	Mary Jones		X			
303 Oak	Annie Jones				X	X

3.1. Using CARRA Kidlink to build AR rosters from the beginning

In this first approach, we used the CARRA Kidlink file to supplement the core AR sources when building the AR rosters on which to train and apply the models. The occupied models were fit using these rosters with Kidlink persons. The Baseline occupied models include, among others, covariates that note whether the AR people are found in a given source at the address (e.g., "On IRS 1040") and whether the AR people are found in a given source at a different address (e.g., "On IRS 1040 elsewhere"). Analogous variables for "On Kidlink" and "On Kidlink elsewhere" were added to the models to note the presence of children from CARRA Kidlink at the given address and also at other addresses. Only addresses that had between one and six people and had a valid household composition were eligible for AR occupied identification. The valid household compositions are one to three adults, with or without children. For more information on the baseline modeling, see Morris et al. (2016) and Mule et al. (2016).

3.2. Adding CARRA Kidlink children to rosters after the AR occupied identification

For this approach, we conducted the AR modeling using only the four core AR sources, without any additional information from the CARRA Kidlink file. This is the Baseline approach, and it identifies a set of AR occupied units. We then added any additional children from the CARRA Kidlink file in the manner described above to the AR occupied units. That is, if any of the AR persons in a unit were listed as a mother or father on the CARRA Kidlink file, then we added all the children from CARRA Kidlink to that unit. For this initial research, we did not place any restrictions on how many children could be added to a unit. Therefore, there may be units that met the household size and composition criteria for AR occupied identification before adding children from CARRA Kidlink, but then did not meet these criteria after the children were added. For this analysis, we still allowed such units to be identified as AR occupied.

3.3. Identifying units to receive the full NRFU contact strategy

For the third approach, we added children to AR occupied units as in the second approach. This uses the Baseline modeling with the four core AR sources to identify a set of AR occupied units. However, here the additional CARRA Kidlink information is used to determine the AR occupied units. Units that are identified as AR occupied by the models but also contain additional children from CARRA Kidlink would be sent to the field to receive the full contact strategy. These are units for which we have an indication of potential undercoverage by the four core AR sources.

4. Results

This section presents the results of the three approaches to using the CARRA Kidlink file in the AR modeling process. To assess the AR modeling approaches, we compare the AR modeling results to the final 2010 Census in terms of occupancy status agreement and household count agreement.

4.1. Using CARRA Kidlink to build AR rosters from the beginning

In this section, we compare the results of the AR occupied identification between the Baseline approach without CARRA Kidlink and the first research approach. For each approach, we identified the top 15 percent of the NRFU universe as AR occupied based on the distance function for the two AR occupied predictive models. See Mule et al. (2016) for more information on the models and the distance function. Note that the individual units identified as AR occupied may differ between the approaches. Table 2

shows the household compositions of the AR occupied units. Overall, the distributions are very similar. The With Kidlink approach identifies slightly fewer units that are one adult with children. Adding children could add some uncertainty to the household roster, which decreases our confidence that the AR roster will match a census response. As a result, fewer of these types of units are identified as AR occupied.

Table 2: Household Compositions of AR Occupied Units

	Baseline		With Kidlink	
Household Composition	Count	Percent	Count	Percent
Total[1]	7,472,483	100.0	7,470,741	100.0
One adult, no children	2,489,150	33.3	2,464,261	33.0
One adult with children	521,024	7.0	431,102	5.8
Two adults, no children	1,944,519	26.0	1,967,713	26.3
Two adults with children	2,101,580	28.1	2,170,219	29.0
Three adults, no children	157,734	2.1	189,268	2.5
Three adults with children	258,476	3.5	248,178	3.3

[1.] Total units differ slightly due to different levels of overlap with AR vacant and AR delete units.

Table 3 shows the status and count agreements for the two approaches overall and by groups defined by the poverty rate for the block group, using the American Community Survey (ACS) five-year estimate for 2006 to 2010. The status agreement shows the percent of the AR occupied cases that were also occupied in the 2010 Census. The count agreement shows the percent of AR occupied cases for which the AR roster and the 2010 Census roster had the same number of people. For these count comparisons, units that were vacant or nonexistent in the 2010 Census had a household count of zero. Overall, using CARRA Kidlink gives a slightly higher count agreement rate (62.5 percent vs. 62.1 percent) as well as a higher rate of units with an AR count higher than census (23.6 percent vs. 22.3 percent). The improvements in the count agreement rate are largest for areas with a higher poverty rate. For example, in the 50 percent or more poverty areas, the count agreement for the With Kidlink approach is 55.2 percent compared to 51.5 percent for the Baseline approach. Note that few of the AR occupied units (less than 2 percent) are in these highest poverty areas. The AR occupied units identified by the With Kidlink approach are slightly more concentrated in low poverty areas.

Table 3: Status and Count Agreement by Poverty Rate

ACS poverty rate of block group	AR Occupied Count	AR Occupied Percent	Occupied Status Agreement	Count Agreement (Row Percent) AR Lower	Agree	AR Higher
Total						
With Kidlink	7,470,741		89.5	13.9	62.5	23.6
Baseline	7,472,483		89.7	15.7	62.1	22.3
0 to 10 percent poverty						
With Kidlink	4,182,862	56.0	90.6	12.9	64.8	22.3
Baseline	4,099,541	54.9	90.8	14.3	64.9	20.8
10 to 20 percent poverty						
With Kidlink	1,850,234	24.8	88.5	14.4	61.0	24.6
Baseline	1,873,478	25.1	88.8	16.2	60.6	23.3
20 to 30 percent poverty						
With Kidlink	803,938	10.8	87.7	15.5	58.8	25.7
Baseline	829,762	11.1	88.1	17.7	57.8	24.5
30 to 40 percent poverty						
With Kidlink	358,681	4.8	87.2	16.8	56.9	26.4
Baseline	375,163	5.0	87.6	19.3	55.3	25.5
40 to 50 percent poverty						
With Kidlink	158,345	2.1	86.9	17.8	55.8	26.4
Baseline	168,396	2.3	87.2	20.9	53.3	25.8
50+ percent poverty						
With Kidlink	116,681	1.6	87.1	19.3	55.2	25.5
Baseline	126,143	1.7	87.5	23.4	51.5	25.1

Table 4 shows the status and count agreement by whether there are CARRA Kidlink persons associated with the unit. The results with CARRA Kidlink have these persons included on the roster. The Baseline results do not have CARRA Kidlink persons added to the roster, but indicate units where CARRA Kidlink persons could be added to the roster based on the rules discussed in Section 3. The rows of the table are defined as follows:

- Has Kidlink-only person: At least one person in the unit is from CARRA Kidlink and this person(s) is not found on the other core sources.
- All Kidlink are new: None of the persons from CARRA Kidlink are found on the other core sources.
- Some Kidlink are new: Some of the persons from CARRA Kidlink are not found on the other core sources, but some other persons from CARRA Kidlink are found on the core sources.
- All Kidlink are corroborated: All of the persons from CARRA Kidlink are found on the other core sources.
- No Kidlink persons: There are no persons from CARRA Kidlink in this unit.

In the With Kidlink approach, 6.8 percent of the AR occupied units have Kidlink-only persons on the roster. These units have a 92.4 percent occupied status agreement rate, 36.7 percent count agreement rate, and a 20.9 percent AR lower count rate. In the Baseline approach, 9.1 percent of the AR occupied units have Kidlink-only persons associated with the unit. Of course, in the baseline approach, those persons are not included in the roster. These units have an indication that there may be children that are not covered by the other core sources. These units have an 89.2 percent occupied status agreement rate. The count agreement is 32.6 percent, and the AR lower rate is 38.6 percent. For these sets of units where there are Kidlink-only persons, the With Kidlink

approach shows improvements in reducing the rate at which the AR count is lower than the census. The two approaches show similar results for units where all CARRA Kidlink persons already exist on the other core sources, as well as for units without any CARRA Kidlink persons.

Table 4: Status and Count Agreement by Presence of Kidlink Persons

Presence of Kidlink-only persons	AR Occupied		Occupied Status Agreement	Count Agreement (Row Percent) AR Lower	Agree	AR Higher
	Count	Percent				
With Kidlink[1]	7,470,741		89.5	13.9	62.5	23.6
Has Kidlink-only person	504,555	6.8	92.4	20.9	36.7	42.4
All Kidlink are new	356,544	4.8	92.2	25.7	39.2	35.1
Some Kidlink are new	148,011	2.0	92.8	9.4	30.6	60.1
All Kidlink are corroborated	1,893,931	25.4	94.2	9.8	69.2	21.0
No Kidlink persons	5,072,255	67.9	87.4	14.8	62.6	22.7
Baseline[2]	7,472,483		89.7	15.7	62.1	22.3
Has Kidlink-only person	677,589	9.1	89.2	38.6	32.6	28.8
All Kidlink are new	325,791	4.4	86.2	31.2	38.2	30.6
Some Kidlink are new	351,798	4.7	92.0	45.6	27.4	27.1
All Kidlink are corroborated	1,967,514	26.3	94.1	10.1	68.4	21.5
No Kidlink persons	4,827,380	64.6	88.0	14.8	63.6	21.6

1. For the With Kidlink approach, the CARRA Kidlink persons are included on the roster and contribute to the household count.
2. For the Baseline approach, the CARRA Kidlink persons do not contribute to the household count. The rows indicate units with a potential undercoverage of children.

Table 5 shows the impact on the population count for NRFU households by age for the Baseline approach and the approach with CARRA Kidlink. The results in this table include the impact of AR vacant and AR delete assignments; these units have a population count of zero. However, the results do not account for the NRFU contact strategy, which includes a single in-person visit to all AR occupied units. Here, the AR result was used for all AR occupied units. For the total population, the Baseline approach had a lower population count than the 2010 Census for adults and young children (those age 0 to 4). With CARRA Kidlink, the AR modeling approach shows a higher count than the 2010 Census for each child age group. The With CARRA Kidlink approach also shows an improvement in the count of adults relative to the Baseline approach. The adult population changes because each approach identifies a different set of units as AR occupied. Recall from Table 2 that the With CARRA Kidlink approach identifies a smaller number of AR occupied units with only one adult.

Table 5. Impact on NRFU Population Count by Age

Age Group	2010 NRFU Count	Baseline Simulation	Baseline Difference	With Kidlink Simulation	With Kidlink Difference
Total	81,496,542	81,138,503	-358,039	81,563,152	66,610
0 to 4	6,906,519	6,868,880	-37,639	6,990,677	84,158
5 to 9	6,610,259	6,656,384	46,125	6,818,970	208,711
10 to 17	9,849,239	9,940,315	91,076	9,997,962	148,723
18+	58,130,525	57,672,924	-457,601	57,755,543	-374,982

Table 6 shows whether the CARRA Kidlink-only persons can be found elsewhere in AR (i.e., at a different MAFID). For the With Kidlink approach, there are 780,832 CARRA Kidlink-only persons in the AR occupied units. The table shows that 16.5 percent of these persons can be found on an IRS 1040 return at a different address. This means that 83.5

percent of these CARRA Kidlink-only persons cannot be found in IRS 1040. This shows a benefit of using CARRA Kidlink to build the household rosters. About 29 percent of the Kidlink-only children are found on CARRA Kidlink elsewhere, meaning that the mother or the father is found at a different address in the core AR sources.

Table 6: Presence of CARRA Kidlink-Only Persons Elsewhere in AR

Presence on other AR sources at a different MAFID	With Kidlink	
	Count	Percent
Kidlink-only persons	780,832	100.0
1040 elsewhere	128,501	16.5
1099 elsewhere	25,305	3.2
Any AR elsewhere (core+VSGI[1])	150,668	19.3
CARRA Kidlink elsewhere	227,609	29.1

1. Veterans Service Group of Illinois Name and Address Resource file

4.2. Adding CARRA Kidlink children to the rosters after the AR occupied identification

We identify the top 15 percent of the NRFU workload as AR occupied, the same 7.5 million units from the Baseline approach presented in Section 4.1. Using Kidlink we added children to these housing units, but we did not fit or score new models. Thus, the total housing units in scope is the same between the Baseline and Kidlink approaches. Table 7 shows the AR household compositions for these 7.5 million units. Of particular interest are the units that do not have any children from the core AR sources but can have children added using the CARRA Kidlink file. The right side of the table shows that adding children from CARRA Kidlink has the greatest relative impact on the one adult with children units. This group increased by 21.5 percent because children were added to units that were previously a single adult without children.

Table 7: Household Composition Prior to and After Addition of CARRA Kidlink

	Baseline		After Kidlink Added		Percent
Household Composition	Count	Percent	Count	Percent	Change
Total	7,472,483	100.0	7,472,483	100.0	
One adult, no children	2,489,150	33.3	2,376,886	31.8	-4.5
One adult with children	521,024	7.0	633,282	8.5	21.5
Two adults, no children	1,944,519	26.0	1,822,740	24.4	-6.3
Two adults with children	2,101,580	28.1	2,223,353	29.8	5.8
Three adults, no children	157,734	2.1	148,567	2.0	-5.8
Three adults with children	258,476	3.5	267,655	3.6	3.6

Table 8 shows the number of units that are affected by the inclusion of CARRA Kidlink children. The top row shows that in 4.4 percent of the AR occupied units, all of the children found on CARRA Kidlink are new to the unit. That is, none of these children from CARRA Kidlink were found in the core AR sources for the unit. In another 4.7 percent of the units, some of the children from CARRA Kidlink are new to the unit while some of the children from CARRA Kidlink were already present in a core AR source.

The third row shows that in 26.3 percent of the units, all of the children from CARRA Kidlink were already present in a core AR source. The CARRA Kidlink information is corroborating these children. In the last row, we see that 64.6 percent of the AR occupied units did not have any children from CARRA Kidlink associated with the unit.

Table 8: Presence of CARRA Kidlink-Only Persons in AR Occupied Units

Presence of Kidlink-only persons	Count	Percent
Total	7,472,483	100.0
All Kidlink are new additions	325,791	4.4
Some Kidlink are new additions	351,798	4.7
All Kidlink are already there	1,967,514	26.3
No Kidlink persons	4,827,380	64.6

Table 9 shows the status agreement for the AR occupied units by the whether CARRA Kidlink children were added to the unit. Overall, about 90 percent of the units were occupied in the 2010 Census. The units in which all of the CARRA Kidlink persons were new additions to the unit have a lower occupied status agreement rate than units in which some or all of the CARRA Kidlink persons were already in the unit. The units in which all of the CARRA Kidlink persons were already in the unit have the highest occupied status agreement. The addition of CARRA Kidlink persons does not impact the status agreement because these are units that were already identified as AR occupied under the Baseline approach. These results show that there is more uncertainty in the housing unit status for units where all of the CARRA Kidlink persons are new additions to the units.

Table 9: Occupied Status Agreement by Presence of Kidlink Persons

	AR Occupied Units		Occupied Status
Presence of Kidlink-only persons	Count	Percent	Agreement
Total	7,472,483	100.0	89.7
Has Kidlink-added person	677,589	9.1	89.2
All Kidlink are new additions	325,791	4.4	86.2
Some Kidlink are new additions	351,798	4.7	92.0
All Kidlink are already there	1,967,514	26.3	94.1
No Kidlink persons	4,827,380	64.6	88.0

Table 10 shows the count agreement by whether CARRA Kidlink children were added to the unit. The table compares the Baseline roster (without CARRA Kidlink) and the roster with CARRA Kidlink children added to the unit. We see that units that have CARRA Kidlink-added persons often have a lower count in AR than the census using the Baseline roster (38.6 percent). When we add the children from Kidlink to these units, this rate of AR undercounts decreases to 10.4 percent, but there is also an increase in the rate of AR overcounts from 28.8 percent to 66.1 percent. This may be an indication that we are erroneously adding children to the unit.

Table 10: Count Agreement by Presence of CARRA Kidlink Persons

	AR occupied		Baseline Roster (Row Percent)			With Kidlink Roster (Row Percent)		
Presence of Kidlink-only persons	Count	Percent	AR Lower	Agree	AR Higher	AR Lower	Agree	AR Higher
Total	7,472,483		15.7	62.1	22.3	13.2	61.2	25.6
Has Kidlink-added person	677,589	9.1	38.6	32.6	28.8	10.4	23.5	66.1
All Kidlink are new additions	325,791	4.4	31.2	38.2	30.6	10.4	16.7	72.9
Some Kidlink are new additions	351,798	4.7	45.6	27.4	27.1	10.5	29.7	59.8
All Kidlink are already there	1,967,514	26.3	10.1	68.4	21.5	10.1	68.4	21.5
No Kidlink persons	4,827,380	64.6	14.8	63.6	21.6	14.8	63.6	21.6

Table 11 shows the impact on the population count for NRFU households by age. Note that this approach does not change the population counts of adults. This approach results

in higher counts of children in all age groups. These large differences over the Baseline approach may again be an indication that this approach erroneously adds many children to the households rosters

Table 11. Impact on NRFU Population Count by Age

	2010 NRFU	Baseline		With Kidlink	
Age Group	Count	Simulation	Difference	Simulation	Difference
0 to 4	6,906,519	6,868,880	-37,639	7,168,961	262,442
5 to 9	6,610,259	6,656,384	46,125	7,015,072	404,813
10 to 17	9,849,239	9,940,315	91,076	10,246,453	397,214

Table 12 shows whether the children that are added to these AR occupied units from CARRA Kidlink are found elsewhere in AR data. One of the motivations for using the CARRA Kidlink data is the concern that there may be children that do not exist on the core AR sources in our AR occupied units. However, our methodology may also be adding children that do exist in the core AR sources at other units. The table shows that over half of the children added from CARRA Kidlink can be found on IRS 1040 at a different address. Similarly, nearly half can be found on CARRA Kidlink elsewhere, meaning that either the mother or father is found on a core AR source elsewhere. There is a large overlap of these children and the IRS 1040 elsewhere children. These results suggest that this approach to using CARRA Kidlink data may be placing many of these children in the wrong unit.

Table 12: Presence of CARRA Kidlink-Only Persons Elsewhere in AR

Presence on other AR sources at a different MAFID	Count	Percent
Kidlink-added children	964,925	100.0
IRS 1040 elsewhere	526,211	54.5
IRS 1099 elsewhere	55,252	5.7
Any AR elsewhere (core+VSGI[1])	546,656	56.7
Kidlink elsewhere	465,033	48.2

1. Veterans Service Group of Illinois Name and Address Resource file

4.3. Identifying units to receive the full NRFU contact strategy

This approach is similar to the second approach (in Section 4.2) in that it uses CARRA Kidlink information after the identification of AR occupied units from the Baseline approach. Here, however, we would not add CARRA Kidlink children to the AR rosters. Instead, any unit with CARRA Kidlink-only children would not be assigned an AR occupied determination. These units would be sent to the field to receive the full contact strategy. This approach acknowledges that there may be undercoverage of children in in the four core AR sources, but also acknowledges that we may not be confident in simply adding all of the Kidlink children to the AR roster as in Section 4.2. By sending these units to the field, this approach lessens the impact on data quality but costs more resources to conduct the additional interviews.

Table 13 shows the count agreement for the initial AR occupied units by whether there are any CARRA Kidlink-only persons in the unit. As seen in the second approach, 9.1 percent of the initial AR occupied units have additional children from CARRA Kidlink. In this approach, we would not call these units AR occupied. Only the 90.9 percent of units without any CARRA Kidlink-only children would receive the AR occupied assignment. The 7.4 million initial AR occupied units is 15 percent of the NRFU workload. The 6.8 million AR occupied units identified by this approach is about 13.6

percent of the NRFU workload. Since more units would need to be visited in the field, this approach would reduce the level of cost savings achieved by the AR modeling.

Units with CARRA Kidlink-only persons show some disagreement both with and without those persons included on the roster. Therefore, it may be beneficial to send these units to the field rather than attempting to add these persons to the AR roster. This approach uses a rule to identify a set of AR occupied units with higher quality while not greatly impacting the number of units identified as AR occupied. If desired, more units could be identified as AR occupied by increasing the threshold for the distance function.

Table 13: Count Agreement by Presence of Kidlink Persons

	Initial AR Occupied Units		Baseline Roster (Row Percent)			With Kidlink Roster (Row Percent)		
Presence of Kidlink-only Persons	Count	Percent	AR Lower	Agree	AR Higher	AR Lower	Agree	AR Higher
Total	7,472,483	100.0	15.7	62.1	22.3	13.2	61.2	25.6
Has Kidlink-only person	677,589	9.1	38.6	32.6	28.8	10.4	23.5	66.1
No Kidlink-only person	6,794,894	90.9	13.4	65.0	21.6	13.4	65.0	21.6

4.4. Summary

The first approach treated CARRA Kidlink as another core AR source to build rosters, train the predictive models on 2010 Census data, and apply the model results. This approach showed small improvements over the Baseline approach in terms of count agreement. The improvements in count agreement were more noticeable in high poverty areas. Only 19 percent of the CARRA Kidlink-only children could be found elsewhere in the core AR sources.

The second approach added children from CARRA Kidlink to units that were identified as AR occupied under the Baseline approach. Adding children from CARRA Kidlink to these rosters resulted in many of the units having AR counts that were higher than the census count. Furthermore, over 50 percent of the children added from CARRA Kidlink could be found elsewhere in the core AR sources. This suggests that this approach may be adding many children to units in which they do not reside.

The third approach took any units from the Baseline approach that had additional children from CARRA Kidlink and made these no determination cases. These units would receive the full NRFU contact strategy. These units had low count agreement rates both with and without CARRA Kidlink, suggesting uncertainty as to which version of the AR roster is best. This approach would reduce the overall AR occupied identification from 15 percent of the NRFU workload to 13.6 percent. Changing the threshold for the distance function would allow more units to be identified as AR occupied, but may add more errors to the census since the quality cutoff would be reduced.

5. Limitations

To assess the AR modeling approaches presented in this paper, we have made comparisons of population counts to the 2010 Census both at the unit level and at aggregated levels. However, the census is not free from errors. In particular, the 2010 Census undercounted children age 0 to 4 and children age 5 to 9 by 4.6 percent and 2.2

percent, respectively (U.S. Census Bureau 2014). Thus approaches that lead to higher counts of children than the 2010 Census may be closer to the truth.

A practical limitation of this simulation is the timing of the data received from the Social Security Administration to build the CARRA Kidlink file. For this research, we used data for all children born up to and including April 1, 2010. The results assume that this data would be present in time for decennial census processing. The AR modeling results must be processed before NRFU begins in mid-May. In a production setting, an older version of the Social Security Administration data may need to be used. This may limit the benefits of the CARRA Kidlink file observed in this research.

6. Conclusions

This paper examined different approaches to adding children from the CARRA Kidlink file in the AR modeling process. The CARRA Kidlink file is a child-to-parent linking dataset. We have examined three methods for using this information to help improve the coverage of children in our core AR sources. This research has shown that the quality of the AR modeling determinations can be improved by incorporating the CARRA Kidlink file in the process. Other factors such as operational complexity, policy, and public perception of these methods must be considered when deciding the best use of the CARRA Kidlink data.

References

Luque, A. and Wagner, D. (2015). “Assessing Coverage and Quality of the 2007 Prototype Census Kidlink Database,” Center for Administrative Records Research and Application (CARRA) Working Paper No. #2015-07, U.S. Census Bureau.

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

Mule, T., Keller, A., Konicki, S., and Kjeldgaard, I. (2016). “The Use of Administrative Records to Reduce Costs in the 2020 Decennial Nonresponse Followup Operation,” Proceedings of the Federal Committee on Statistical Methodology Policy Conference, December 2016.

Wagner, D. and Layne, M. (2014). “The Person Identification Validation System (PVS),” Center for Administrative Records Research and Application (CARRA) Working Paper No. #2014-01, U.S. Census Bureau.

U.S. Census Bureau (2014). “The Undercount of Young Children,” Census Bureau Working Paper.