Using Administrative Records to Reduce Census Nonresponse Followup Operations

Vincent Thomas Mule, Jr and Andrew Keller
U.S. Census Bureau

Abstract

The Bureau is examining ways to reduce the costs of the 2020 Census while maintaining a high quality census. One of the largest contributors to the 2010 Census cost was when enumerators had to be sent to interview and resolve housing units that did not return a mail questionnaire in the Nonresponse Followup operation. One area of research has been attempting to utilize administrative record information from various federal and other sources to reduce the number of interviews to be done for the Nonresponse Followup operation. This paper will present an overview of the current research being conducted and some initial research of possible administrative records usages.

Introduction

As part of the Census 2020 planning, the U.S. Census Bureau has been examining ways that we can reduce costs while still maintaining our commitment to quality. One area being researched is the possibility of using administrative record information to reduce the dependence on in-person visits for data collection. To investigate possible alternatives to reduce the Nonresponse Followup (NRFU) workload for 2020, the Census Bureau has been researching using administrative records as a mechanism for resolving units in lieu of NRFU visits by FRs. This paper presents an overview of some research being done to determine the possibility of using administrative records in this capacity during the NRFU operation.

2010 Nonresponse Followup Operation

In Census 2010, the Census Bureau conducted the NRFU operation to resolve addresses after a self-response was unable to identify them as occupied, vacant, or non-existent. NRFU consisted of up to six contacts by Census Bureau field representatives (FRs). After which, remaining unresolved units were imputed a status and household size via count imputation. The NRFU universe consisted of about fifty million addresses.

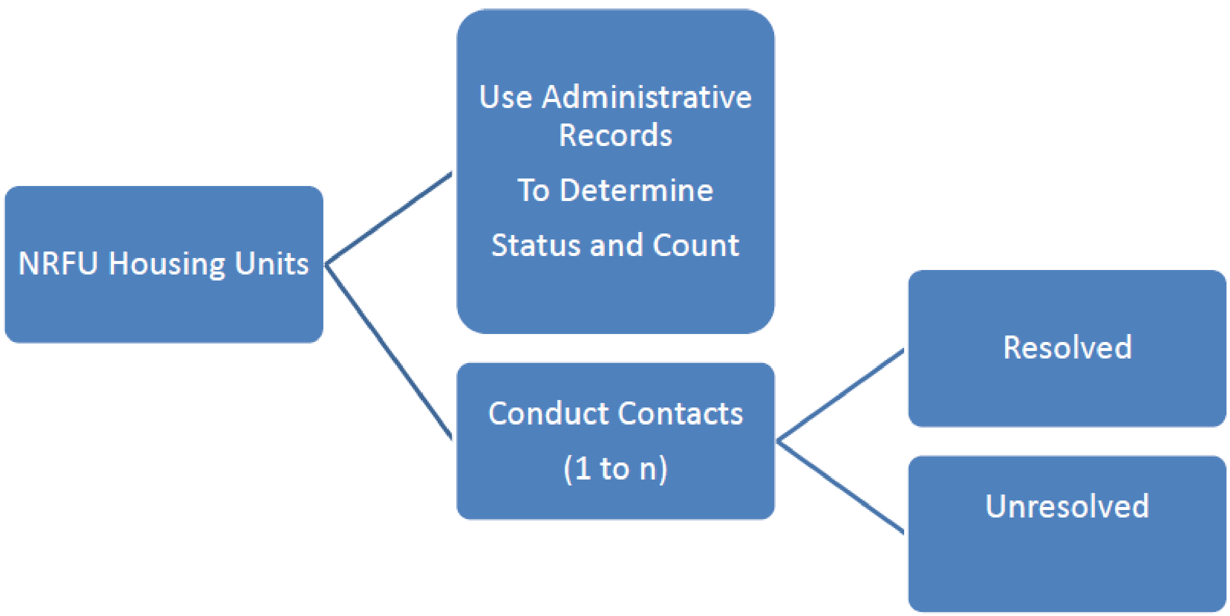
The 2010 Nonresponse Followup Operation occurred from mid-April until the middle of August and was composed of four field operations: Nonresponse Followup (NRFU), Nonresponse Followup Reinterview (NRFU RI), Nonresponse Followup Vacant Delete Check (VDC), and Nonresponse Followup Residual (Residual). The bulk of the 2010 Nonresponse Followup Operation was concentrated on NRFU. NRFU RI was designed as a quality check on NRFU enumerators' results. VDC verified vacant and delete housing units from the NRFU operation as well as a first time enumeration of some units that were not in the original NRFU universe. Residual (1) obtained population counts for units that were known to be occupied but where the count was unknown and (2) first enumerations that were not in the original NRFU or VDC universes. See Jackson et. al (2012) for more information on the 2010 Nonresponse Followup Operation.

Administrative Records Usage and Contact Reduction Approaches

To reduce the NRFU workload, the Census Bureau has been researching using administrative records as a mechanism for resolving units in lieu of completing field visits. Two central questions are *when and how* administrative records should be used. With respect to *when*, should administrative records be used before any NRFU visits are completed or after a suitable number of NRFU visits is completed. With respect to *how*, the question is whether administrative records should be used to assign only unoccupied status. This sends all possible occupied cases into the NRFU universe. Or, is it also possible to also assign an occupied status and population count via administrative records. We outline three possible approaches.

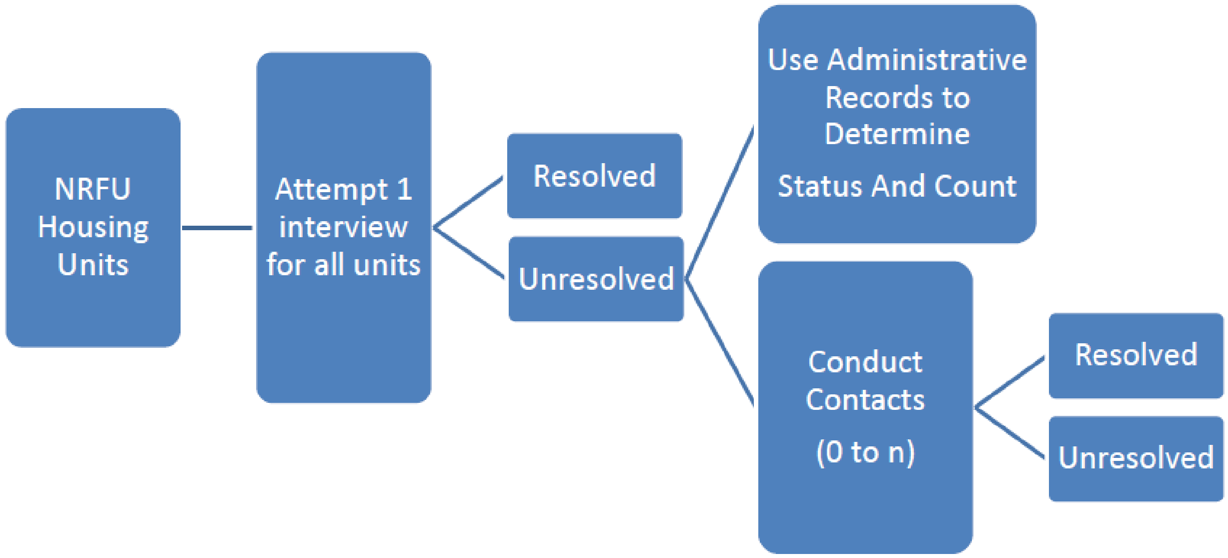
This first approach is to use administrative records first. That is, suppose we allow for the use of administrative records to resolve the entire NRFU universe before any NRFU visits are completed. We use administrative records, when available and of sufficient quality, to assign a status (occupied, vacant, or non-existent) and a population count if occupied. Then, we conduct NRFU operations for the remaining cases. Figure 1 shows a flow chart of this approach.

Figure 1: Administrative Records First



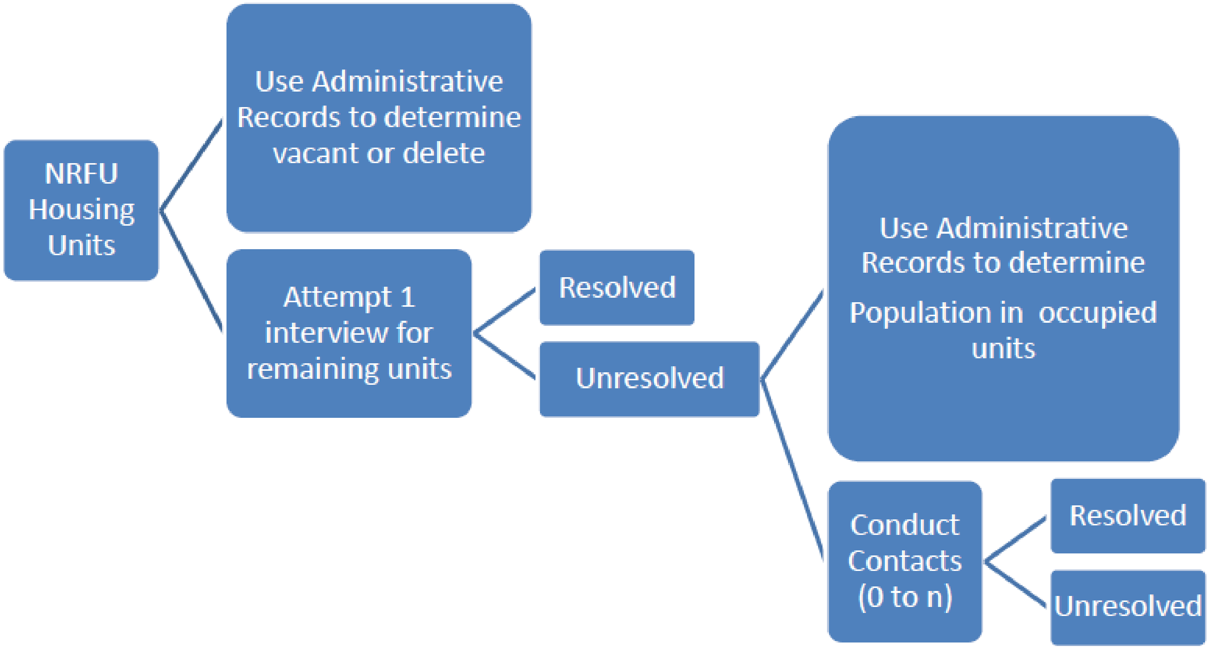
This second approach is to use administrative records after first allowing a visit attempt at every NRFU unit. For cases that remain unresolved, we allow for the use of administrative records to resolve the remaining NRFU universe. Again, we use administrative records to assign a status (occupied, vacant, or non-existent) and a population count if occupied. Then, we conduct NRFU operations for the remaining cases. Figure 2 shows a flow chart of this approach.

Figure 2: NRFU Visits First, Administrative Records After One Visit



Last, suppose we try a hybrid ordering of administrative records and NRFU visits. To do this, suppose we first allow for the use of administrative records only to resolve whether units as unoccupied only (vacant or non-existent). Doing this helps reduce the number of trips by FRs to units where no one can be located to complete the interview. Next, we allow a visit attempt at every remaining NRFU unit. This allows an interview to be completed by an enumerator. It also allows the household to self-respond if they go online using the link provided in the notice of visit.At this point, we use administrative records to assign an occupied status and population count. Then, we conduct NRFU operations for the remaining cases. Figure 3 shows a flow chart of this approach.

Figure 3: Hybrid Administrative Records and NRFU



Issues and Challenges with Using Administrative Records Instead of Contacting

In this section, we will go over some of the issues and challenges of using administrative records instead of in-person contact visits. These are the five main issues and challenges being faced with this research. In this section, we will expand on each in more detail.

- Which sources do you use?
- If a person is associated with numerous addresses, which do you use?
- Which person records do you use for a NRFU unit?
- How do you make the Administrative Records usage versus Contact decision?
- What demographic characteristics do you use for Administrative Record people?

Which sources do you use?

There are several administrative record sources available for use. These include federal, state and third party sources. There are Federal sources like the Internal Revenue Service (IRS) Individual Income Tax 1040 Returns and the Center for Medicare and Medicaid Services Medicare Database. There are additional federal files maintained by other agencies for administering those programs that could be sources of information. There are programs like the Supplemental Nutrition Assistance Program (SNAP) that are administered at the state level. The Census Bureau is researching using information like this that could be obtained by the fifty states and the District of Columbia.

A third set are Third Party sources. These are commercial databases that have person and address information. Examples of these include Experian, Targus and InfoUSA. Some of these third party sources have been seen in the Census Match Study to not have a large coverage of the child population.

In using administrative records, one decision is which of the federal, state and Third Party sources do we want to utilize. The Administrative Records modeling team is examining the quality implications of using these different sources. We do realize that in addition to the quality implications there are also possible policy concerns especially with the use of Third Party files and how that data was collected by the different vendors.

If a person is associated with numerous addresses, which do you use?

One possibility with using multiple administrative record sources is that a person might be associated with multiple addresses. This causes a potential issue for census enumeration if we want to associate that person with only one address. Even with a single source file, the same person can be found to be associated with different housing units. One example is that a person may be receiving SNAP benefits in February at one address and then receiving SNAP benefits in March at a different address.

The Administrative Records Modeling team has been examining ways to associate a person with an address. One way has been to build models based on which address was the administrative record counted during the 2010 Census. Logistic regression models have been run to see if we can predict which address is more likely to be where the person was counted based on the characteristics of the person and their address. These models take into account census information about the housing unit from our Master Address File but also administrative record information that we might have about the person. A second approach has been to look at transaction or record dates on the different records. A rule-based approach can use the record with the transaction that is closer to Census Day of April 1^{st}.

Which person records do you use for a NRFU unit?

Another challenge is to decide which person records to use in place of doing an in-person NRFU visit. One example is that you have 102 Main Street. You have an IRS 1040 tax filing for 102 Main Street that was processed in February. You have also received SNAP information from the state that benefits were given to the family living at 102 Main Street in March. In this instance, we have two sources of information about the people associated with 102 Main Street. Figure 4 shows a possible Venn diagram of the potential overlap of the people on these two sources.

Figure 4: Potential Overlap of Person Information from Multiple sources.

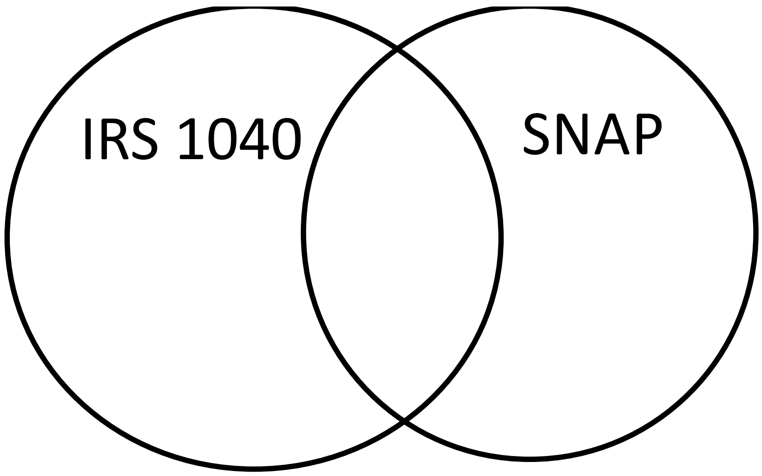


Figure 4 shows that you can have three possible sets of people for 102 Main Street from both of these sources:

- listed on both the IRS 1040 return and SNAP application.
- Listed on only the IRS 1040 return
- Listed on only the SNAP application.

In utilizing these two administrative record sources for 102 Main Street, the challenge is to determine which person records to use. There are several possibilities that include:

- Use all of the records from both sources
- Use only the person records from the IRS 1040 return
- Use only the person records from the SNAP return
- Use the people reported on both the IRS 1040 and the SNAP application.
- Other combination of the two sources

In this scenario, the two sources of information have some but not all of the people in common. There is another scenario where the two different sources are providing two entirely different households. The IRS 1040 return can be providing the Smith Family and the SNAP application can be providing the Jones family. In this instance, we need to determine which to use for Census Day April 1st.

How do you make the Administrative Records usage versus Contact decision?

With the determination of which administrative records to associate with the NRFU housing unit, the next point is to decide about whether you want use these administrative records or to continue to do in-person field visits. In the Hybrid scenario, one example would be deciding whether to make the unit occupied with those administrative record people after one unsuccessful field visit or to conduct more field work.

The Administrative Records Modeling team has been examining rule-based and predictive models as possible ways to make the decision. One example of rule-based decisions is to use information based on a reporting of IRS and United States Postal Service (USPS) information. The team has examined predictive models about whether the housing unit count is correct or if the administrative record sources are putting the person in the same place as the 2010 Census. These predictive models produce probabilities that are correlated with having higher agreement with Census counts. Predictive models have been developed using logistic regression, decision tree and random forest approaches. Morris (2014) documents some initial work using logistic regression, decision trees or random forest modeling approaches. Her work is a continuation of earlier work done by David Brown looking at logistic regression approaches.

What demographic characteristics do you use for Administrative Record people?

Some administrative record sources like the IRS 1040 return has name information but does not have characteristics usually obtained during the census enumeration like age, sex, race or Hispanic Origin. One of the challenges with utilizing administrative records during the enumeration is to figure out how to assign characteristics to these people.

For characteristics like age and sex, the vast majority of administrative records have a Social Security Number or Individual Tax Identification Number. These can be linked to the Social Security Numident or other sources files that have age, date of birth and sex information. A further challenge is to obtain the race and Hispanic origin of these people. One possibility is that the person might have been enumerated in the 2010 Census. If so, the race and Hispanic origin can be obtained from a past Census Bureau response. If this is not available, then other administrative record source files may have information collected about the race and Hispanic origin of the person. This can be used in a rule-based approach or the other source characteristics can be utilized in an imputation model. One potential reason for using an imputation model is that the administrative record sources might not have the level of detail that has been reported by the Census Bureau in the past. Examples of this detail include being able to report multiple races or being able to report detailed levels like Mexican for Hispanic origin or Japanese for Asian.

Count and Characteristic Imputation Research

While researching ways to use administrative records in place of in-person visits and the impact of potentially reducing the number of contacts, this may result in more housing units having an unresolved status than in the past. In Census 2000 and Census 2010, it was less than 0.5% of the time when the status of the unit and population count could not be determined. For census purposes, the status is whether the unit is a) occupied, b) vacant or c) does not meet our definition of a housing unit and should be removed from the list.

With reduced contacts, the amount of housing units having an unresolved determination will probably be of larger magnitudes than seen in the past. With this possibility, the Administrative Records Modeling team is researching ways to compensate for this increase. The team’s work has focused on ways of modifying the approach used for count imputation in the 2010 census. One possibility is to utilize more of a log-linear model approach that utilizes covariates based on administrative records in the imputation.

With this reduction, there may also be an increase in the amount of demographic characteristics like age, sex, race and Hispanic origin that are missing. The Administrative Records Modeling team is examining several ways to try to compensate for this during the imputation process. The Center for Statistical Research Methods (CSRM) staff on the project are developing software that can implement log-linear imputation approaches to fill in missing characteristics. Griffin (2014) also examines using non-ignorable approaches to account for missing characteristics like Hispanic origin.

Conclusion

This paper provides a summary of some of the initial ideas of using administrative records to potential reduce the amount of in-person data collection necessary in the 2020 Census. It documents the issues and challenges of using administrative records in this capacity. It also documents some initial research for imputation for missing housing unit and person information based on the usage of administrative records or the reductions in contacts.

References

Griffin, R. (2014), "Issues concerning imputation of Hispanic origin due to administrative record enumeration in the 2020 Census" In the2014 Joint Statistical Meetings Proceedings for the Survey Research Methods Section,. .

Jackson, G. et al (2012) "2010 Census Nonresponse Followup Operations Assessment Report," 2010 Census Planning Series Number 190.

Morris, D. (2014), "A Comparison of Methodologies for Classification of Administrative Records Quality for Census Enumeration" In the2014 Joint Statistical Meetings Proceedings for the Survey Research Methods Section,.