2020 Census

2018 End-to-End Census Test Study Plan: Administrative Record Usage for Nonresponse Followup

Draft Pending Final Census Bureau Executive Review and Clearance.

Version 1.2

08/15/2017

[This page intentionally left blank]

Contents

1. Introduction....4
1.1 Purpose of Study....4
1.2 Background....5
2. Assumptions....7
3. Methodology....7
4. Data Requirements....8
5. Division Responsibilities....9
6. Analysis Milestone Schedule....9
7. Risks/Limitations....9
7.1 General Program Risks for All Evaluations, Experiments, or Assessments (That Also Apply to This Evaluation, Experiment, or Assessment)....9
7.2 Risks Specific to This Evaluation, Experiment, or Assessment....9
7.3 Limitations (Known at This Time)....10
8. Administrative Record Usage Analysis....10
8.1 NRFU Administrative Record Mailings....10
8.2 Household-Level Comparison Analysis....11
9. References....12
10. Documentation....12
10.1 Sensitivity Assessment....12
10.2 Review/Approval....13
11. Version History....13
12. Glossary of Acronyms....14

1. Introduction

To meet the strategic goals and objectives for the 2020 Census, the Census Bureau must make fundamental changes to the design, implementation, and management of the decennial census. These changes must build upon the successes and address the challenges of previous censuses while also balancing challenges of cost containment, quality, flexibility, innovation, and disciplined and transparent acquisition decisions and processes.

For the Nonresponse Followup (NRFU) operation, the 2016 Census Test allowed us to obtain more information about these three questions related to administrative record usage:

1. What can we learn about how the United States Postal Service determines and assigns mail pieces as being Undeliverable-as-Addressed (UAA)?
2. How many additional responses are received for administrative record vacant, delete, and occupied units based on the additional NRFU mailing?
3. How do the administrative record vacant, delete, and occupied determinations compare with fieldwork results?

In the 2018 End-to-End Census Test, we will investigate similar questions to those from the 2016 Census Test related to administrative record usage. We will also learn if our modified contact strategy improves vacant, delete, and occupied match rates:

1. How many additional responses are received for administrative record vacant, delete, and occupied units based on the additional NRFU mailing?
2. How do the administrative record vacant, delete, and occupied determinations compare with fieldwork results?
3. Does the requirement of UAA on the administrative record mailing improve administrative record vacant and delete match rates to fieldwork determinations?
4. Does the requirement of the absence of UAA on the administrative record mailing improve administrative record occupied match rates to fieldwork determinations?

1.1 Purpose of Study

The goal of this analysis is to continue to evaluate the use of administrative records to reduce the number of visits in the NRFU workload. Administrative record information will be used in predictive models to reduce the number of contacts during the NRFU operation by identifying administrative record vacant, delete, and occupied housing units. We identify occupied housing units using several administrative records sources: Internal Revenue Service (IRS) individual taxpayer returns (IRS 1040), IRS informational returns (IRS 1099), Indian Health Services (IHS) Patient Database, Social Security Administration Numident file, and Center for Medicare and Medicaid Services Medicare Enrollment Database (CMS MEDB). These sources are used to build a household roster for addresses and as covariates in the predictive models. A third-party source is also used. The third-party source is used as a covariate to indicate if the person was observed in administrative record sources either at this or another address. The Fitness for Use team provided information about past decennial census responses and best determinations of race and Hispanic origin from the decennial census and other sources.

Rastogi and O'Hara (2012) describe these data sources in the Census Match Study. The IRS 1040, IRS 1099, and MEDB are data from other federal agencies. For the 2016 Census Test, we also used UAA information from the United States Postal Service (USPS) from the first and second mailing attempts.[1] We implemented predictive modeling approaches to identify vacant and delete units in the 2016 Census Test, and we will repeat this in the 2018 End-to-End Census Test.

1.2 Background

The 2018 End-to-End Census Test is an important opportunity for the Census Bureau to ensure an accurate count of the nation's increasingly diverse and rapidly growing population. It is the first opportunity to apply much of what has been learned from census tests conducted throughout the decade in preparation for the nation's once-a-decade population and housing census. The 2018 End-to-End Census Test will be held in three locations, covering more than 700,000 housing units: Pierce County, Washington; Providence, Rhode Island; and the Bluefield-Beckley-Oak Hill, West Virginia, area.

The 2018 End-to-End Census Test will be a dress rehearsal for most 2020 Census operations, procedures, systems, and field infrastructure to ensure there is proper integration and conformance with functional and nonfunctional requirements. The test also will produce a prototype of geographic and data products. Note that the 2018 End-to-End Census Test results are based on three sites that were purposely selected and cannot be generalized to the entire United States.

The NRFU fieldwork will be implemented with the addition of the Enterprise Census and Survey Enabling (ECaSE) platform in this test. This test will use the administrative record approach revised based on lessons learned from the 2016 Census Test.

In the 2016 Census Test, the fieldwork resolution for administrative records units showed that for administrative records vacant (AR vacant) and administrative records delete (AR delete) units in our evaluation sample, 21.1 percent and 29.1 percent, respectively, were occupied. Table 1 has detailed results regarding this resolution. This proportion of fieldwork occupied units was higher than expected. To guard against this result in the 2018 End-to-End Census Test, the administrative records modeling team has planned for a revised contact strategy.

[1] The USPS classifies mail that cannot be delivered by postal mail carriers as UAA, and such mail is sent into a special operation.

Table 1: NRFU Fieldwork Resolution Compared With AR Prediction

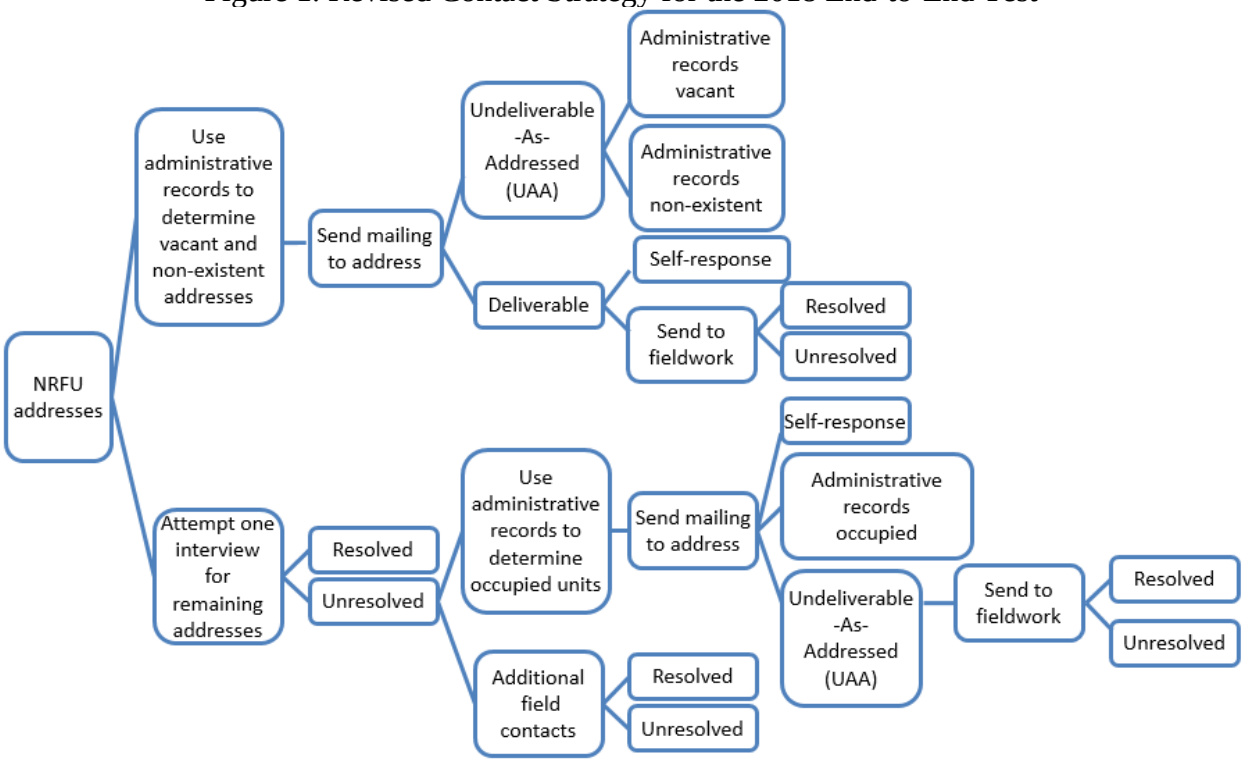
	Total	Occupied		Vacant		Delete		Unresolved	
	N	%	SE	%	SE	%	SE	%	SE
AR Occupied, Phase 1									
Total	2,338	80.3	0.7	4.8	0.4	1.7	0.2	13.2	0.6
LA County	1,241	87.3	0.9	3.4	0.5	1.2	0.3	8.1	0.7
Harris County	1,097	72.4	1.2	6.4	0.7	2.3	0.4	19.0	1.0
AR Vacant									
Total	715	21.1	1.3	42.8	1.6	20.7	1.2	15.4	1.2
LA County	236	19.9	2.3	43.2	2.9	22.5	2.3	14.4	2.1
Harris County	479	21.7	1.6	42.6	2.0	19.8	1.5	15.9	1.4
AR Delete									
Total	313	29.1	2.1	10.9	1.4	48.6	2.2	11.5	1.7
LA County	172	24.4	2.9	7.6	1.9	57.0	3.3	11.0	2.2
Harris County	141	34.8	3.2	14.9	2.1	38.3	2.7	12.1	2.5
AR Occupied, Phase 2									
Total	64	51.6	5.9	4.7	2.2	1.6	1.4	42.2	5.7
LA County	35	57.1	8.0	5.7	3.7	2.9	2.6	34.3	7.6
Harris County	29	44.8	8.9	3.4	2.2	0.0	0.0	51.7	8.6

Figure 1 shows the revised contact strategy for the 2018 End-to-End Census Test. Administrative record vacant and delete units will be identified before the start of NRFU and removed from fieldwork. Units identified with good administrative record occupied information will receive one visit from an enumerator during NRFU before becoming eligible for an administrative record enumeration. For the nonadministrative record cases, a maximum of six contact attempts will be made. Starting on the third visit, enumerators will be allowed to resolve cases using proxy interviews if they are unable to contact a household member.

For administrative record cases, an additional mailing will be sent during the NRFU operation to all cases. The 2016 Census Test analysis showed that only 57.3 percent of the AR vacant and 75.4 percent of the AR delete cases had this additional mailing marked as UAA by the USPS. For the units for which this additional mailing could be delivered, the lack of UAA information potentially conflicts with the initial AR determination of vacant or delete. As part of the new contact strategy in the 2018 End-to-End Census Test, the UAA information from this additional mailing will be used to make final decisions about the administrative record determinations. If the administrative record determination is vacant or delete and the mailing comes back with UAA information, the determination will remain. Otherwise, the unit will go to the field for the

full contact strategy (i.e., up to six contact attempts). We anticipate that this new requirement of UAA information for the additional mailing will improve the rate at which AR vacant and delete units match the fieldwork results. A similar change will be made for the AR occupied units. If the administrative record determination is occupied and the additional mailing does not come back as UAA, then the determination will remain. Otherwise, the unit will go to the field for full contact strategy. In the 2016 Census Test, only 1.5 percent of the AR occupied cases were UAA on the additional mailing. The contact strategy for the 2016 Census Test did not update the initial AR determinations based on the UAA status of the additional mailing.

Figure 1. Revised Contact Strategy for the 2018 End-to-End Test



For the administrative record identification, there will be two phases of identification during the NRFU period. Similar to the 2016 Census Test, the first phase will happen about four days before the start of NRFU. The NRFU-eligible units will be processed using the predictive approaches to identify the administrative record vacant, delete, and occupied units. Approximately the first week of June, a new delivery of IRS 1040 data will be made available to the Census Bureau and the second phase of occupied identification will happen. As in the 2016 Census Test, we will conduct an additional set of processing when this new source, an updated list of IRS 1040 returns, is available to be able to identify additional administrative record occupied units for which contacts can be reduced.

2. Assumptions

This study has one major underlying assumption: The administrative record data available for the 2018 End-to-End Census Test is consistent with what will be available for the 2020 Census in terms of content and timing.

3. Methodology

This section lays out the methodologies that will be used in conducting the administrative record usage analysis with 2018 End-to-End Census Test data.

The Concurrent Analysis and Estimation System (CAES), based on specifications from the Administrative Records Modeling team, will be using predictive model approaches to identify

vacant, delete, and occupied units. These approaches were also implemented in the 2016 Census Test. For the identification of vacant units, we will use results from the 2010 Census NRFU operation to build a predictive model. Examples of variables in the model include Undeliverable-as-Addressed information from the United States Postal Service for the 2018 End-to-End Census Test mailings, presence of person records on administrative record and third-party files, information about the address from the Master Address File, and American Community Survey (ACS) 5-year estimates for the block group[2]. This model will be used to predict the probability of a unit being occupied, vacant, or delete. These predicted probabilities will be used to identify addresses that are more likely to be vacant and less likely to be occupied. Similarly, the probabilities will be used to identify addresses that are more likely to be delete and less likely to be occupied.

For occupied units, the Administrative Record Modeling team has researched two models to identify occupied units. Household rosters were built using information IRS 1040, IRS 1099, Center for Medicare and Medicaid Services Medicare Enrollment Database, and the Indian Health Services Patient Database. A person-place model was developed to predict the likelihood that a person was correctly associated with an address. Morris (2014) documents how this approach developed predictive models using person-matching results. This work has shown to be correlated with count agreement between administrative records and the decennial census. We have also developed predictive models for household compositions. These compositions are based on the number of adults and number of children (if any) present in the unit. This work has been able to predict the decennial census household composition that would be observed given the administrative record composition and other information. Morris et al. (2016) document the methodology used in the 2015 Census Test. Like the 2016 Census Test, the 2018 End-to-End Census Test will use a distance function approach instead of linear optimization (which was used in the 2015 Census Test) to identify the AR occupied units using information from both of these model predictions. The distance function approach, documented in Kjeldgaard and Konicki (2017), uses both the person-place probability and the household-composition probability. If the combination is above a predetermined threshold, a unit will qualify for administrative record enumeration.

4. Data Requirements

We require the following files to conduct this research study:

- IRS tax year 2017, 2016, 2009, and 2008 1040 file.
- MEDB 2017, 2016, and 2009 files.
- Indian Health Service 2017, 2016, and 2009 files.
- IRS tax year 2017 and 2009 information return 1099 file.
- 2018, 2017, and 2010 Numident File.
- 2018, 2017, and 2016 American Community Survey MAF extracts.
- 2017 and 2009 VSGI Files.
- 2018 End-to-End Census Test and 2010 Census UAA results file.
- Delivery Point Validation File.

[2] Examples of ACS estimates include the poverty rate, mobility rate, and vacancy rate at the block group level

- 2018 End-to-End Census Test housing unit and person results .
- 2018 End-to-End Census Test postprocessing file.
- 2006-2010 and 2011-2015 American Community Survey (ACS) 5-year estimates.

5. Division Responsibilities

Division	Responsibility
Decennial Statistical Studies Division	Project management, statistical analysis, report writing
Decennial Census Management Division	Project management, budget
Center for Administrative Records Research and Applications	Assign PIKs to 2018 End-to-End Census Test postprocessing file

6. Analysis Milestone Schedule

Activity	Start	End
Prepare initial draft of study plan	05/17	07/17
Present initial draft of study plan to DROM	06/17	06/17
Prepare final draft of study plan	06/17	07/17
Develop SAS programs	01/18	11/18
Data processing	06/18	11/18
Administrative Record Usage Analysis	06/18	11/18
Prepare initial draft of Administrative Record Usage Analysis Report	06/18	12/18
Present initial draft of report to DROM	01/19	01/19
Prepare final draft of report	01/19	04/19

7. Risks/Limitations

7.1 General Program Risks for All Evaluations, Experiments, or Assessments (That Also Apply to This Evaluation, Experiment, or Assessment)

- Funding difficulties may cause redirection of staff resulting in project not being completed on time.

7.2 Risks Specific to This Evaluation, Experiment, or Assessment

- Not receiving UAA reason codes for the 2018 End-to-End Census Test mailings. If we do not receive UAA reasons for the mailings, then we will not be able to determine if certain codes are better tied to a vacant or delete outcome. Without UAA reasons, we are unable

to implement a Kappa statistic predictor in our model, as it assesses the rate of agreement in the UAA codes between successive mailings.

7.3 Limitations (Known at This Time)

- The assumption listed in Section 2 may not be met.

8. Administrative Record Usage Analysis

The goal of this study is to evaluate the use of administrative records to remove cases from the NRFU workload.

We want to learn how administrative records can best be implemented in the 2020 Census. We will complete an analysis evaluating the methodology that was used in the 2016 Census Test. To do this evaluation, we will assess the accuracy with which occupied, vacant, and delete cases were identified by the administrative records modeling. Because there is no control panel for the 2018 End-to-End Census Test, we will select a 1-in-5 sample of administrative record occupied, vacant, and delete cases that will still go out to the field. This will allow the housing-unit-level comparisons done in past tests to be repeated in this test. These comparisons will allow us to assess the impact of the new contact strategy on the accuracy of the administrative record determinations.

8.1 NRFU Administrative Record Mailings

In this test, an additional postcard mailing will be sent to the administrative record vacant, delete, and occupied units, including those units that we sample to go out to the field. One of the things that we want to quantify is the additional number of self-responses that were received after these mailings were sent to households. A household could self-respond by using the internet, calling the Census Questionnaire Assistance (CQA) phone number, or returning their paper questionnaire. For this test, we will quantify how often self-responses were received after the NRFU mailing occurred.

We will also quantify the portion of the initially identified administrative record vacant and delete units that received UAA on the additional postcard as well as the portion of the initially identified administrative record occupied units that did not receive UAA on the additional postcard.

By sending the additional postcard to all administrative records cases, we can quantify the utility of incorporating UAA information from the final mailing in the final administrative record determination of vacant and delete. We will analyze how often administrative record vacant cases resolved to vacant in the field by whether they were associated with UAA on the final mailing. Similarly, we will analyze how often administrative record delete cases resolved to delete in the field by whether they were associated with UAA on the final mailing. In the 2016 Census Test, the AR determination cases in the evaluation sample were treated as regular NRFU fieldwork cases and thus did not receive the additional mailing. Sending the additional mailing to all administrative records cases in the 2018 End-to-End Census Test will allow us to assess

whether the new contact strategy improves the rate of agreement between administrative records and fieldwork results.

8.2 Household-Level Comparison Analysis

Based on sending a sample of cases out for fieldwork, we can compare the NRFU interview results against the administrative record results for units that were determined to be occupied, vacant, or delete by administrative records. We anticipate that the new contact strategy for the 2018 End-to-End Census Test will improve the rate for which the administrative record determinations agree with fieldwork results, especially for vacant and delete units.

This analysis will be divided into three parts: analyzing addresses determined by our model to be vacant, analyzing addresses determined by our model to be delete, and analyzing addresses determined by our model to be occupied. For each part, the analysis will help to show the impact of the new contact strategy for this test.

<u>Analyzing addresses that our approach predicted would be vacant:</u>

1. For cases that were predicted to be vacant but sent to fieldwork, compare against NRFU interview status of address (occupied, vacant, delete, unresolved).
2. For cases that were predicted to be vacant with a vacant NRFU interview status, develop a list of reasons as to why they were not identified by the administrative records modeling.
3. For cases that were predicted to be vacant, compare against NRFU interview status of address by whether or not the unit received UAA on the additional postcard mailing.
4. For cases that were predicted to be vacant but resolved to a NRFU status of occupied, analyze characteristics of people, housing units, and UAA, and analyze the National Change of Address (NCOA) file to investigate if there was an indication that someone moved out of the home before Census Day or into it after Census Day.

<u>Analyzing addresses that our approach predicted would be delete:</u>

1. For cases that were predicted to be delete but sent to fieldwork, compare against NRFU interview status of address (occupied, vacant, delete, unresolved).
2. For cases that were not predicted to be delete with a delete NRFU interview status, develop a list of reasons as to why they were not identified by the administrative records modeling.
3. For cases that were predicted to be delete, compare against NRFU interview status of address by whether or not the unit received UAA on the additional postcard mailing.
4. For cases that were predicted to be delete but resolved to a NRFU status of occupied, analyze characteristics of people, housing units and UAA, and analyze the National Change of Address (NCOA) file to investigate if there was an indication that someone moved out of the home before Census Day or into it after Census Day.

<u>Analyzing addresses that our approach predicted would be occupied:</u>

1. For cases that were predicted to be occupied but sent to fieldwork, compare the AR counts against the NRFU interview population count.

2. For cases that were predicted to be occupied, compare against NRFU interview status of address (occupied, vacant, delete, unresolved).
3. For cases that were not predicted to be occupied with an occupied NRFU Interview Status, develop a list of reasons as to why they were not identified by the administrative records modeling.

9. References

Kjeldgaard, I., and Konicki, S. (2017), "Results of Administrative Record Use in the 2016 Census Test."

Morris, D.S. (2014), "A Comparison of Methodologies for Classification of Administrative Records Quality for Census Enumeration," in JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 1729-1743.

Morris, D.S., Keller, A., and Clark B. (2016). "An Approach for Using Administrative Records to Reduce Contacts in the 2020 Census," Statistical Journal of the International Association of Official Statistics, 32 (2016): 177-188.

Rastogi, S. and O'Hara, A. (2012) "2010 Census Match Study."

10. Documentation

The following sections document the sensitivity assessment that this study plan has undergone and the list of individuals who have reviewed and approved it.

10.1 Sensitivity Assessment

This table specifies whether or not the document contains any administratively restricted information.

Verification of Document Content	
This document does not contain any: • Title 5, Title 13, Title 26, or Title 42 protected information; • Procurement information; • Budgetary information; and/or, • Personally identifiable information.	
Document Author/Team Lead:	Date:

10.2 Review/Approval

This table documents the review level and/or approval authority.

Document Review and Approval Tier: Program Document		
Name	**Area Represented**	**Date**
Tom Mule	Team Lead	
Maryann Chapin	2020 Program Manager	
Patricia McGuire	Systems Engineering and Integration Manager	
Kimberly Higginbotham	Program Management Manager	
Deb Stempowski	Chief, Decennial Census Management Office	

11. Version History

The document version history recorded in this section provides the revision number, the version number, the date it was issued, and a brief description of the changes since the previous release. Baseline releases are also noted.

Rev #	Version	Date	Description
0	1	5/18/17	Initial Draft
1	1	7/6/17	Comments from DROM and QP Reviewer
1	2	8/2/17	Approved by DROM

12. Glossary of Acronyms

Acronym	Definition
ACS	American Community Survey
CAES	Concurrent Analysis and Estimation System
CARRA	Center for Administrative Records Research and Applications
CMS MEDB	Center for Medicare and Medicaid Services Medicare Enrollment Database

Acronym	Definition
CQA	Census Questionnaire Assistance
ECaSE	Enterprise Census and Survey Enabling
IHS	Indian Health Services
IRS	Internal Revenue Service
IRS 1040	Internal Revenue Service individual taxpayer returns
IRS 1099	Internal Revenue Service informational returns
MAF	Master Address File
NCOA	National Change of Address
NRFU	Nonresponse Followup
PIK	Personal Identification Key
UAA	Undeliverable-as-Addressed
USPS	United States Postal Service
VSGI	Veterans Services Group of Illinois