United States Census 2020

2020 Research and Testing:
2013 Census Test Assessment

United States™ Census Bureau

Version 2.1

May 28, 2014

[This page intentionally left blank]

Contents

List of Tables

[This page intentionally left blank]

Executive Summary

The 2013 Census Test was an exploratory study of new methods for data collection in the 2020 Census nonresponse follow-up (NRFU) operation. The impetus for the test was an assessment of 2020 research and testing activities in early 2013 that highlighted the need for research on possible census cost savings, including interviewer productivity (i.e., cases completed per hour) in NRFU. The aim of the 2013 Census Test was to pilot several novel methods that have potential to reduce NRFU costs.

The methods explored in the 2013 Census Test include: 1) using administrative records to reduce the NRFU workload; 2) reducing the number of enumeration contact attempts; 3) using an adaptive case management strategy to control personal visit enumeration effort; and 4) using telephone to make initial enumeration contact attempts. These methods were tested in an experimental design that permitted comparisons of different kinds of administrative record use, number of contacts, case management, and telephone contacts.

The test simulated NRFU conditions by attempting to enumerate addresses whose occupants did not self-respond to the 2010 Census. There was no self-response phase in the study. Rather, sample units were notified by mail that they were selected to participate. The self-response phase was excluded to save time and money and to focus on NRFU methods.

Some 2,077 housing units participated in the study. Six pairs of block groups were selected from a universe of block groups in Philadelphia, and one block group in each pair was randomly assigned to an adaptive case management or a "fixed" strategy, resulting in a total of twelve block groups divided into two groups.

The adaptive group employed a computer-assisted telephone interviewing (CATI) operation to call numbers matched to sample addresses for up to two weeks, after which cases were transferred to in-person interviewers. Computer-assisted personal interviewing (CAPI) interviewers in the adaptive group were provided high priority cases on a daily basis, determined by response propensity models, and instructed to attempt these high priority cases each day they worked.

In the fixed group, CAPI interviewers were instructed to make two calls on different days and at different times to each telephone number attached to their assigned cases prior to making personal visits to the addresses. Interviewers in the fixed group were given a set of cases that were not prioritized and trained to perform personal visit attempts to any of these cases after performing telephone calls on all cases with numbers.

All sample housing units were randomly assigned to two groups - those in which records, if available, were used to remove cases from the workload before going to the field and those in which administrative records, if available, were not used to remove cases. Cases left in the group where records were used to remove cases received a maximum of three personal visit attempts.

Cases in the other group received a maximum of one or three personal visit attempts, depending on the treatment.

This design resulted in four treatments:

- Treatment One:
  - Administrative records used to remove cases from workload.
  - CAPI interviewers called addresses with telephone numbers before visiting.
  - No cases were high priority.
  - Three personal visits attempted before seeking a proxy respondent.
- Treatment Two:
  - Administrative records not used to remove cases from workload.
  - CAPI interviewers called addresses with telephone numbers before visiting.
  - No cases were high priority.
  - Three personal visits attempted before seeking a proxy respondent.
- Treatment Three:
  - Administrative records used to remove cases from workload.
  - CATI interviewers called addresses with telephone numbers before sending to CAPI interviewers.
  - Propensity models determined high priority cases.
  - Three personal visits attempted before seeking a proxy respondent.
- Treatment Three:
  - Administrative records not used to remove cases from workload.
  - CATI interviewers called addresses with telephone numbers before sending to CAPI interviewers.
  - Propensity models determined high priority cases.

If administrative record could enumerate case, one personal visit attempted before seeking a proxy respondent. If not, three personal visits attempted.

The test provided an initial opportunity to use administrative records to inform removing cases from the NRFU workload. To assign a housing unit as vacant, it was necessary to receive Undeliverable-as-Addressed information from the United States Postal Service obtained from the prenotice letter mailing. We assigned a vacant status correctly for 67.9% (19/28) and 80.3% (49/61) of the cases in Treatments 2 and 4. For enumeration of occupied housing units, at least one of three administrative records sources needed to have the same validated persons within a source across two years. We found 42.2% (73/173) of the housing units had a household size based on administrative records that was smaller, and 22.5% (39/173) had a household size that was larger than that found in the NRFU interview.

The test examined a reduction in the number of maximum personal visit attempts. In total 345 of 1,642 cases were incomplete (i.e., stopped in the field due to maximum attempts). Of these, 100 could be enumerated with administrative records while 245 had no data.

The test required CAPI interviewers follow strict contact procedures. Over 85 percent of transmissions were compliant on days interviewers worked, occurring between 6:00 a.m. and midnight. Of cases on their final personal visit attempt, 28.46% were attempted only one time and left uncompleted. Implementation failures make it more difficult to judge propensity score predictions and experimental manipulations. Both supervisors and interviewers said that more exercises relating to contact strategies would have helped them understand unique procedures in the test.

Commercial vendors performed the matching of phone numbers to addresses for this study, and the .quality of matching was problematic. Only 2.5% of fixed and 4.4% of adaptive cases that had telephone numbers and were not removed with administrative records were completed by telephone in CATI or, for fixed cases, by telephone prior to any personal visit attempts being made. Two-fifths of telephone contact attempts by in-person interviewers resulted in no telephone contact. Over one-third of call attempts from the CATI facility did not result in telephone contact, and over half of the telephone numbers sent to CATI were ruled ineligible based on non-contact.

Treatments where administrative records were removed before fieldwork had lower productivity than other treatments because easier cases were removed, and the cases left were more dispersed. However, interviewers charged fewer hours in treatments where administrative records were removed before fieldwork, leading to cost savings albeit lower productivity. An adaptive approach to case management also resulted in a jump in productivity The results suggest that, even with the difficulties previously described, the use of administrative records to removed workload and the implementation of an adaptive design case management approach have potential to reduce costs.

Completion rates for cases not removed before fieldwork were comparable between the fixed (77.8%) and adaptive (80.8%) groups, as were completion rates that counted all cases whose status could be determined with administrative records as a complete (86.2% for fixed and 90.1% for adaptive).

Assessment of the 2013 Census Test suggests that the four novel methods listed previously can be implemented in difficult field conditions. It appears that reducing the field workload with administrative records and using an adaptive case management approach can lead to cost savings.

The research revealed a number of operational issues that can inform the conduct of future 2020 research and testing activities. Among the issues noted are several recommendations:

- **Research different sources and criteria for inferring from administrative record and postal information the occupancy status and population size of housing units.**
- **Research methods for obtaining proxy interviews after maximum contacts have been made on sample units.**

- **Assess the use of administrative records and methods of imputation for cases with no data after interviewing is complete.**
- **Measure the quality of proxy interviews performed on occupied housing units;**
- **Make case management and interviewing systems more "user friendly" for enumerators.**
- **Research ways to achieve higher enumerator compliance with case management instructions and interviewing procedures including ways to tailor training, supervision, and automation.**
- **Research alternative methods for determining the response propensity of sample units on the next interviewer attempt.**
- **Research how to make daily case assignments for enumerators, including the size of daily workload and its geographic composition.**
- **Explore alternative ways to match telephone numbers with sample addresses and rank order them for likelihood of contact.**
- **Explore alternative methods of determining the quality of telephone numbers matched to addresses.**
- **Examine alternative calling algorithms for CATI efforts to enumerate sample households.**
- **Examine the tradeoffs and approach to "telephone first" procedures in the field.**
- **Standardize the metrics for telephone call outcomes**
- **Develop cost and progress reports suited to complex experimental designs**

The test was accomplished with existing Census Bureau systems, data resources, equipment, and personnel. A small group of headquarters staff members teamed with several administrators in the Philadelphia Regional Office and 18 field interviewers to carry out the research. The test was authorized in June 2013, and data collection was completed in the first week of December 2013. (Data collection was delayed two weeks by the failure of appropriations in October 2013.)

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 the previous Censuses while also balancing challenges of cost containment, quality, flexibility, innovation, and disciplined and transparent acquisition decisions and processes.

The 2013 Census Test piloted novel uses of administrative records and contact strategies to reduce the nonresponse follow-up (NRFU) workload and increase NRFU efficiency. The test was not designed to make or infer comparisons across treatments or to create statistical estimates; it primarily tested the operational efficiency of innovative procedures applied for the first time in the field.

1.1 Scope

The results of the 2013 Census Test will inform the use of administrative records and NRFU contact strategies tested during the 2020 Research and Testing Program and reduce the risks associated with a larger scale implementation of these methods, which are planned for the 2014 Census Test.

1.2 Intended Audience

This report is intended for the program managers and staff responsible for researching and implementing the 2014 and 2015 Census Tests, contractors supporting the 2020 Census, and oversight organizations.

2. Background

Decennial Census field costs for the NRFU operation in 2010 totaled $1,589,397,886 (Walker et al., 2012). To control costs, the Census Bureau is testing strategies to decrease NRFU workload and conduct a more efficient data collection operation (Business Plan for the 2020 Census in Support of the FY 2013 Budget Submission, 2012).

The 2013 Census Test grew out of discussions in the spring of 2013 on how to focus the 2020 research and testing effort on attacking the major cost drivers in the census. Efficiencies in nonresponse follow-up are essential in order to meet 2020 budget targets. The methods examined in this test are targeted at achieving a more efficient NRFU process. These methods have not been examined in previous census tests.

The test explored operational aspects of workload reduction and data collection. The primary goal of the test was to assess whether the Census Bureau could implement a simulated NRFU

data collection using different contact strategies and administrative records during production. Thus, the primary goal was to test operational feasibility.

The test examined:

1. decreasing NRFU workload using administrative records to identify vacant housing units and enumerate occupied housing units, and
2. achieving more efficient data collection (i.e., interviewer productivity) by
   a. lowering the cap on computer-assisted personal interviewing (CAPI) contact attempts,
   b. using telephone - either a centralized computer-assisted telephone interviewing (CATI) operation or calls made by individual enumerators - to make initial contacts for cases whose addresses can be matched to telephone numbers, and
   c. using an adaptive design approach to manage case assignments.

Background on these procedures as they relate to cost and data quality is provided below.

2.1 Using Administrative Records to Reduce NRFU Workload

The Census Bureau is committed to designing and conducting a 2020 Census that costs no more per housing unit than the 2010 Census, while continuing our commitment to maintain high quality results. A major cost driver for the 2010 Census involved collecting information from housing units that did not self-respond. To reduce costs for the 2020 Census, the Census Bureau is investigating the strategic reuse of federal, state, and private data sources. The 2010 Census Match Study evaluated data from federal agencies and commercial vendors, finding that administrative records data were useful for determining housing unit and occupancy status, informing household population count, and assigning demographic characteristics. Using a composite of many commercial and federal files, the study found that administrative records were present for 89.2 percent of NRFU housing units in the 2010 Census and for 79.5 percent of persons enumerated in NRFU units in the 2010 Census (Rastogi and O'Hara, 2012).

The Census Bureau is evaluating administrative records sources in various combinations to enhance data collection and processing methods for the NRFU operation. Different administrative records files and Census Bureau data are used to explore agreement and disagreement observed in content across sources to develop approaches (such as modeling or business rules) to improve NRFU data quality and reduce costs through fewer in-person visits.

The Census Bureau also uses administrative records data to simulate NRFU operation results if fieldwork were curtailed or reduced and to evaluate the impact of reduced field operations on the vacant and delete check operations used in previous censuses. Administrative records are evaluated to help determine if and when to interview a housing unit and to improve the contact strategies for non-responding housing units.

For the 2020 Census, the Census Bureau is investigating methods to decrease the NRFU workload. In the 2013 Census Test, we looked at the use of record information in a small-scale field setting. Sample addresses were matched to records from various sources. In two experimental treatments, record information was employed to reduce NRFU workload by removing some vacant housing units and "enumerating" some cases. We flagged occupied housing units using three administrative records sources: IRS Individual Taxpayer Returns (IRS 1040), Center for Medicare and Medicaid Services Medicare Enrollment Database (CMS MEDB) and Targus Federal Consumer (TFC) data. Rastogi and O'Hara (2012) describe these three data sources in the Census Match Study. The IRS 1040 and MEDB were data from other federal agencies and the TFC data were obtained from the commercial vendor Targus. For the 2013 Census Test, we also used Undeliverable-as-Addressed (UAA) information from the United States Postal Service (USPS) obtained from the prenotice letter mailing.[1] We flagged vacant housing units if the prenotice mailing associated with the 2013 Census Test resulted in an UAA designation with a "vacant" reason code.

Information from records in another treatment was used to prioritize field effort. In this case, sample addresses received one personal visit if adequate record information was available to enumerate households. Sample units without adequate record information received up to three in-person visits. This adaptive design case management approach stands in contrast to a "fixed" contact strategy, which did not specify different numbers of contact attempts for different addresses.

2.2 Reducing Maximum Personal Visit Attempts to Increase NRFU Efficiency

In the 2010 Census, production salary cost and mileage contributed 63.9 percent and 22.7 percent to the total cost of the NRFU field operation, respectively. Field staff made 104,432,798 total contacts attempts during the operation. Enumerators were permitted up to six contact attempts for each case. About 41 percent of the 47,197,405 housing units included in the NRFU operation were enumerated on the first contact, and nearly 83 percent were enumerated in three or fewer contact attempts (Walker et al., 2012).

The 2013 Census Test used a maximum of three personal visit contact attempts. In one treatment, the presence or absence of administrative records suitable to enumerate the household determined the level of contact effort. For cases with record information available, one contact attempt was made to enumerate the household. For cases without record information, up to three contact attempts were permitted.

2.3 Implementing an Adaptive Design Case Management Strategy to Increase NRFU Efficiency

An adaptive survey design involves a tailored, dynamic approach to managing cases, enabled through a centralized data collection system. This approach utilizes paradata, process

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

information such as contact history and case dispositions, response data, and auxiliary frame information to guide contact procedures and allocate resources (Miller, 2013).

Adaptive survey designs are being tested to increase data collection efficiency while maintaining data quality. For example, the National Survey of College Graduates, conducted by the U.S. Census Bureau on behalf of the National Science Foundation, included a subsample of 4,000 cases in its 2013 adaptive design experiment. These cases were monitored throughout the course of data collection, and cases were switched to different modes or to active or inactive status to decrease nonresponse bias or to reduce cost (Coffey et al., 2013). The National Survey of Family Growth monitors subgroup representation in the sample during data collection in order to adjust field efforts to attack differential nonresponse (Groves and Heeringa, 2006).

The 2013 Census Test used an adaptive approach to case management to tailor effort expended on each case (using record information as described above) and to alert field representatives (referred to in this report as interviewers) to the cases most likely to be completed with data (i.e., occupied, vacant, or completed using the telephone questionnaire assistance) on the next contact attempt. A program containing a model with parameters from administrative record information, auxiliary frame data, and paradata produced response propensity scores on a daily basis. Case management systems highlighted each interviewer's seven highest propensity-to-respond cases, and interviewers were trained to attempt these "high priority" cases each day they worked.

The adaptive case management approach was compared to the fixed approach, in which in-person interviewers were provided with all of their cases at the beginning of the field period and instructed to use their best judgment in making contacts with sample units, which was similar to the approach used in the 2010 NRFU operation. In the fixed approach, we did not provide interviewers with a set of high priority cases to contact each day.

2.4 Employing Telephone Calls Before In-person Follow-up to Increase NRFU Efficiency

A number of Census Bureau surveys, including the American Community Survey (ACS), are conducted in multiple modes, beginning with the least expensive - a mailed paper questionnaires (and now Internet), followed by phone interviews, and then personal visits. This practice orders the least expensive interview methods at the beginning, and interview attempts get progressively more expensive as subsequent modes are employed to reach nonrespondents. Following this logic, the 2013 Census Test examined the use of telephone prior to personal visits. Two different telephone methods were studied – a centralized CATI approach and decentralized telephone calls made by individual interviewers prior to attempting personal visits. These different approaches model ways in which the Census Bureau could organize telephone data collection in the 2020 Census.

3. Methodology

3.1 Site Selection and Sample

For this study, 2,077 sample addresses were selected from a universe of pre-specified block groups in the Philadelphia Metropolitan Statistical Area. Philadelphia was chosen due to Census CAPI interviewer availability and operational and cost advantages provided by working with a Census regional office close to Census Bureau headquarters. To save time and money, the test did not contain a self-response component. Only housing units that did not mail back a self-response form during the 2010 decennial census were eligible for inclusion. Thus, the NRFU data collection environment was simulated using information from the 2010 Census.

Block groups with more than five percent of the population in group quarters were removed during sample selection because the research focused on individual housing units. Block groups with more than 10 percent of households having no one over age 14 that speaks English "very well" or better were also removed because of limited availability of CAPI interviewers to conduct interviews in non-English languages.

From the remaining block groups, an iterative process was followed in which one block group was randomly selected. A similarity score between this block group and all other eligible block groups was calculated to produce pairs of block groups, one to be assigned to the adaptive case management group and the other to the fixed contact group. The similarity score was based on a weighted combination of the absolute differences between block groups for the following ten block group variables, found in the Census Bureau's 2012 Planning Database (2012 Planning Database, 2013), that relate to likelihood of contacting the household (Groves & Couper, 1998):

1. Percent vacant
2. Average number of persons per household
3. Percent owner occupied
4. Percent single unit
5. Percent multi-unit (ten or more)
6. Percent NRFU 2010 cases
7. Percent Black
8. Percent Hispanic
9. Percent under age 18
10. Percent over age 64

Each randomly sampled block group was paired with the block group with the highest similarity score. If the similarity score was above a threshold, this pair was included in the sample and removed from the remaining block group universe. This block group sampling strategy was repeated until there were eight pairs of block groups. Two matched block groups were discarded due to anticipated logistical difficulties noted by the Philadelphia Regional Office, leaving six matched block groups from which sample housing units were subsequently selected.

Treatments were then assigned to housing units using a two-step process. One block group in the matched pair was assigned randomly to the fixed contact strategy and the other to the adaptive design contact strategy, taking care to balance sample size between the two treatments. All sample housing units were then randomly assigned to two groups - those in which records, if available, were used to remove cases from the workload before going to the field and those in which administrative records, if available, were not used to remove cases. Nine housing units that had recently participated in other Census surveys were removed from the sample before this random assignment. This process resulted in 2,077 eligible housing units across four treatments.

3.2 Study Design

3.2.1 Study Design - Administrative Records

Two treatments (Treatments 1 and 3 in Table 1) employed administrative records to remove from field workload the occupied housing units that had records deemed suitable to enumerate them. Sample unit occupancy was estimated by United States Postal Service (USPS) information on whether advance letters, mailed two weeks prior to data collection, were deliverable or not, and by examination of administrative records information. The suitability of records for enumerating these housing units was determined through the Census Bureau's research on matching administrative records information to 2010 Census NRFU housing units. We classified sample units as vacant if prenotice letters mailed to them were returned with Undeliverable-as-Addressed "vacant" reason code from the USPS and if they had no other record evidence of occupancy from the administrative record sources that were used. See section 3.3.2 for more details. In Treatments 1 and 3, these vacant housing units were also removed from the field workload.

The Census Bureau did not employ administrative records to reduce workload in Treatments 2 and 4. Instead, administrative records prioritized the effort to be expended on cases in Treatment 4, and there was no use of records in Treatment 2.

Table 1. Study Design for 2013 Census Test

	Fixed	Adaptive
Admin. records remove vacant and occupied units before fieldwork	**Treatment 1** Administrative records • Remove cases from workload Telephone • If number, CAPI interviewers call • All numbers called twice Priority • *None* Number of visits • Three personal visits before proxy	**Treatment 3** Administrative records • Remove cases from workload Telephone • If number, CATI before field • CATI call procedures Priority • Propensity models determine high priority cases Number of visits • Three personal visits before proxy
Admin records do <u>not</u> remove vacant and occupied units before fieldwork	**Treatment 2** Administrative records • *Not used* Telephone • If number, CAPI interviewers call • All numbers called twice Priority • *None* Number of visits • Three personal visits before proxy	**Treatment 4** Administrative records • Determine level of effort (i.e., number of contacts) Telephone • If number, CATI before field • CATI call procedures Priority • Propensity models determine high priority cases Number of visits • If administrative record, one personal visit before proxy • Three personal visits before proxy

3.2.2 Study Design - Telephone

Prior to data collection, the Census Bureau attempted to match all sampled housing units to telephone numbers using purchased vendor files. Up to three numbers were matched to each sample address. Some 70 percent of sample addresses were linked to at least one number. The numbers included both landline and cell. The Census Bureau verified all landline telephone numbers to make sure they were in service.

In the fixed contact strategy treatments (Treatments 1 and 2), the Census Bureau instructed individual CAPI interviewers to make two calls on different days and at different times to each telephone number attached to their assigned cases prior to making personal visits to the addresses. Interviewers then attempted up to three in-person contact attempts for sample units not reached by telephone and the sample units without matched telephone numbers. If a CAPI interviewer could not complete an interview after three in-person contact attempts, they were instructed to obtain a proxy interview from a knowledgeable source such as a neighbor.[2]

In the adaptive case management group (Treatments 3 and 4), the Census Bureau employed a CATI operation to call numbers matched to sample addresses. Interviewers at the Census Bureau's Tucson CATI facility, the Tucson Contact Center (TCC), attempted interviews for up to two weeks. Nonresponding CATI cases were transferred to in-person interviewers. Housing units without telephone numbers were sent directly to CAPI interviewers at the beginning of the field period.

All in-person interviewers, whether in the fixed or adaptive group, used laptops with computer-assisted personal interviewing (CAPI) software for case management and questionnaire administration. The laptops permitted us to send daily case priorities to CAPI interviewers in the adaptive case management group and permitted all interviewers to transmit completed cases daily.

3.2.3 Study Design - High Priority Cases

CAPI interviewers in the adaptive group (Treatments 3 and 4) were provided high priority cases on a daily basis, determined by response propensity models. Inputs to propensity models included auxiliary frame information and paradata. Paradata were obtained from the Contact History Instrument (CHI), which CAPI interviewers were instructed to fill out after each contact attempt. Auxiliary frame information came from three sources:

1. government or other administrative records containing such information as household size and age composition,
2. 2010 decennial contact history information used to determine a housing unit's initial propensity to respond, and
3. other frame information such as whether or not the housing unit was located in a multi-unit building.

3.3 Use of Administrative Records

For each of the three administrative records sources, IRS Individual Taxpayer Returns (IRS 1040), Medicare Enrollment Database (MEDB), and Targus Federal Consumer (TFC), we obtained 2011 and 2012 files. For IRS 1040, these were for the tax years 2011 and 2012, which

[2] A proxy interview is conducted with someone who has knowledge of the address but does not live at the household, such as a neighbor, relative, or building manager.

were created in 2012 and 2013, respectively. For TFC, there was a file available for each quarter of 2012, but we used the fourth quarter file only. We converted each file into a person-level file where a unique combination of Master Address File Identification number (MAFID) and Protected Identification Key (PIK) define a person. PIK assignment means that the person record was given an unidentifiable unique person identifier. See Wagner and Layne (2012) for more information on PIK assignment. A PIK may be duplicated across multiple MAFIDs. We retained all records where the MAFID and PIK were non-blank. For the MEDB files, we removed from consideration any records that had a non-blank beneficiary date of death. We found all units for which there were unvalidated person records. A PIK could not be assigned because there was insufficient information to uniquely identify them for these records. Units with any unassigned PIKs were not allowed to be flagged as occupied.

3.3.1 Occupied Housing Unit Flag

We applied two rules to flag housing units as occupied. These rules were applied separately for each of the three record sources, so a housing unit may have "passed" for some sources but failed for others. If a housing unit passed the rules for at least one file, it was flagged as occupied. The two rules were:

- For the given source, the housing unit must have the same set of PIKs (i.e., persons) on the 2011 and 2012 version of the source. That is, each PIK in the unit on the 2011 file must be in the unit on the 2012 file and vice versa.
- There must not be any unvalidated person records for the housing unit on either year of the given administrative records source. That is, each person in the housing unit needed to be assigned a PIK.

Any unit that passed these two rules for a given administrative records source was flagged as occupied due to that source. All persons in the given administrative records source were used to generate a household population count and define certain characteristics of the housing unit. A unit could also be flagged due to multiple sources. These sources may have identified different sets of persons. All of the persons from all sources were retained. If these persons overlapped between multiple sources, they were only counted once in the housing unit.

3.3.2 Vacant Housing Unit Flag

For the test, we used UAA information from the United States Postal Service following the test prenotice mailing. Units that were UAA with a "vacant" reason code were flagged as vacant.

3.3.3 Forming the Universe

We integrated the use of administrative records into three phases. Each sample unit was eligible to receive initial occupied and vacant flags, regardless of treatment even though Treatments 2

and 4 did not use these flags to remove cases before going to the field. A flag was set for an occupied or vacant unit as defined above. Table 2 shows the initial flags by treatment.

Table 2. Initial Flags by Treatment

Treatment Description	Overall Total	Initial Occupied Flag	Initial Vacant Flag
1 – Fixed, removal before field	511	162	42
2 – Fixed, no removal before field	510	153	34
3 – Adaptive, removal before field	528	174	50
4 – Adaptive, no removal before field	528	155	74
Total Units	**2,077**	**644**	**200**

In the second phase, we removed both flags for cases with both an initial occupied and vacant flag. The rationale behind the removal was that inconsistent information existed between administrative record sources and UAA information. We wanted to send these to the field to resolve the conflicting information. In total, there were sixteen cases with both the initial occupied and initial vacant flags set. After we removed the flag, there were 628 cases with a final occupied flag and 184 with a final vacant flag. Table 3 adds two additional columns onto Table 2, indicating cases with a final occupied and vacant flag by treatment.

Table 3. Final Flags by Treatment

Treatment	Overall Total	Initial Occupied Flag	Initial Vacant Flag	Final Occupied Flag	Final Vacant Flag
1	511	162	42	160	40
2	510	153	34	150	31
3	528	174	50	166	42
4	528	155	74	152	71
Total	**2,077**	**644**	**200**	**628**	**184**

In the third phase, we determined the cases that would be removed from the workload and those used to prioritize field effort and for post-study analysis. Since Treatments 1 and 3 use enumeration via administrative records, we removed those cases from post-study analysis because field operations were not done to determine their final housing status and population count. Consequently, we perform our analysis on cases in Treatments 2 and 4. Table 4 adds four additional columns onto Table 3, indicating the treatment membership of the removed and analysis cases.

Table 4. Removed Cases by Treatment

Treat.	Overall Total	Initial Occupied Flag	Initial Vacant Flag	Final Occupied Flag	Final Vacant Flag	Occupied Removed	Vacant Removed	Occupied Analysis	Vacant Analysis
1	511	162	42	160	40	160	40	0	0
2	510	153	34	150	31	0	0	150	31
3	528	174	50	166	42	166	42	0	0
4	528	155	74	152	71	0	0	152	71
Total	**2,077**	**644**	**200**	**628**	**184**	**326**	**82**	**302**	**102**

In Section 5.1, administrative record analysis is focused on Treatments 2 and 4.

3.4 Adaptive Case Management Methodology

Adaptive case management relied on the execution of a SAS program. This program contained code that counted contact attempts and prioritized cases using a discrete-time hazard model (a type of logistic regression model). The program also executed a series of business rules that communicated cases to be removed from workload, which cases were high priorities, and which cases were on their last contact attempts to the case management.

3.4.1 Case Prioritization Code

Cases' response propensity scores were based on two logistic regression models that used auxiliary frame information and paradata in the field. The "Day 0 Model" was run once, generating propensity scores for cases with no field paradata. The second "Case Prioritization Model" was run nightly. This second model used paradata obtained daily from the Census Bureau's Universal Tracking System (UTS). UTS is a centralized repository for paradata and cost and progress data for a number of Census Bureau surveys. Its purpose is to track survey operations including case workloads and different modes. In addition to obtaining paradata, UTS served as the production environment to run the program (data setups, model, and business rule execution).

The Day 0 Model used 2010 decennial Census information from the Lifecycle Analysis Team (LCAT) data file to predict a household's likelihood of response in the 2013 Census Test. Initial work was completed to determine which groups of variables were appropriate for the model. These variables included housing unit status variables, refusal indicators, respondent information, and household characteristics.

Three main-effects stepwise models were then run on 2010 NRFU cases in the Philadelphia Metropolitan Statistical Area to determine which variables were significant in predicting likelihood of response at the first, second, and third personal visit. Some manual examination and variable changes were made to increase model parsimony. Due to the high predictive value

of the main-effect models, 2-way interactions were excluded.[3] These data were then split into two panels, and parameters were calculated on one panel and scored on the other panel to test the model.

It was determined that the first personal visit model (as opposed to models that predicted response likelihood at the second or third personal visit) was most appropriate because we wanted to predict the likelihood of a completed response on Day 1, or at the first contact. Table 5 lists the final parameters used in the Day 0 Model. This model was then used to score all cases in the 2013 Census Test sample. Because case assignment occurred after Day 0 priorities were assigned, the top third propensity scores were assigned "high priority" status on the first day of data collection in the field.

For subsequent days in the field, a discrete-time hazard model estimated the propensity to obtain a completed interview on the next contact attempt. The logistic model modeled the probability, $p_{ic,}$ of obtaining an interview for case, *i*, on each contact attempt, *c* via:

$log\left(\frac{p_{ic}}{1-p_{ic}}\right) = \alpha + \beta_1 x_{ic1} + \cdots + \beta_k x_{ick}$, where $x_{ic1}, \ldots, x_{ick}$ are the covariates on the next contact attempt, *c*, for the i^{th} case associated with that contact, and $\beta_1 + \cdots + \beta_k$ are the regression parameter estimated from all prior contact attempts.

The dependent variable predicted whether a contact resulted in a completion, with five outcomes qualifying as completions: occupied, sufficient partial, vacant, temporarily occupied, and telephone questionnaire assistance (TQA) complete. The model included twelve predictor variables. See Table 6 for the description and location of each variable. The four variables derived from administrative record sources were chosen because they relate to likelihood of contacting a type of household (Groves & Couper, 1998). Contact records recorded for proxy respondents were not included in the model.[4]

The Case Prioritization Model ran in UTS each morning around 3:00 a.m. Eastern Standard Time after data from CAPI interviewer laptops and the Regional Office Survey Control System (ROSCO) were pulled into UTS tables. UTS housed one "fixed frame" file and two tables used by the program. The fixed frame file included study design information, information from the master address file, and information from administrative records including whether or not an address could be enumerated with records and information on cases with records like whether or not all sample unit members were less than 30 years-old. The tables included a daily snapshot of

[3] Predictive value was determined using concordance, how often the model correctly predicted that a response occurred within the 1st, 2nd, or 3rd visit.

[4] This is because traditionally positive predictors of contact and cooperation for the sampled housing unit (e.g., in-person contact attempts and no prior reluctance) are generally negative predictors of proxy contact attempts, which happens when sampled housing units cannot be reached or refuse to participate in the survey.

all CAPI cases in ROSCO, the “Household Table,” and a snapshot of all contact attempts recorded in the field, the “Contact History Table.”

Table 5. Day 0 Model Parameters

Variable	Definition	Valid Values
CENSTAT2010_ACSX	2010 Census Status (Final Tabulation, Listed on MAF)	0 - Housing Unit Not in 2010 Census 1 - Housing Unit in 2010 Census - Respondent Form 2 - Housing Unit in 2010 Census - Enumerator Form 3 - Group Quarters Not In 2010 Census - Vacant 4 - Group Quarters Not In 2010 Census - Non-Existent 5 - Group Quarters Not In 2010 Census - Unknown Status 6 - Group Quarters Not In 2010 Census - Duplicate 7 - Group Quarters Not In 2010 Census - Non-Residential
DSFSPR10	Delivery Sequence File Status (Listed on MAF)	0 - Not on the indicated DSF 1 - Flagged as residential on the indicated DSF 2 - Flagged as commercial on the indicated DSF 3 - Flagged as an 'X' record on the indicated DSF
PP_HOUSING_STATUS_32	Housing Status File (CUF)	1 - Occupied 2 - Vacant 3 - Delete 5 - Unresolved
HUTYP	Housing Unit Type (Final Tabulation, Listed on the MAF)	M - Multi-unit O - Other S - Single Unit T - Trailer Null - Not captured
lang32_recode	Recode of Languages Used During Interview	Various Categories
final_action_ac	MAF Action Code from Address Canvassing	A - Add K - Change not involving address C - Change involving address M - Move (Add/Delete block move) Q - Add in combination with another action (not

		identified as a Move) L - Duplicate D - Delete N - Nonresidential R - Rejected from MTDB processing
REF_32	NRFU Refusal Indicator	0 – No 1 - Yes
RESPONDENT_32	NRFU Respondent Type	0 – Missing 1 – Household Member 2 – Proxy 99 - Multiple
TENURE_32	NRFU Tenure Type	0 - Missing 1 - Owned with Loan 2 - Owned without Loan 3 - Rent 4 - No payment of rent 99 - Multi
UNITSTAT_FTX	Unit Status Code from MAF Extract Created in 2010	01 - Valid Living Quarters 02 - Demolished 04 - Nonexistent 05 - Provisional Add 06 - Under Construction 07 - Duplicate 08 - Empty Mobile Home/Trailer site 11 - Unable to Locate 29 - Physical Merge 31 - Other uninhabitable
trailer_enumx	Enumeration Trailer Flag	Y – Unit is a potential trailer N – Unit is not a potential trailer
ubsa_range_enumx	Basic Street Address Size	1 - 1 unit 2 - 2 to 4 units 3 - 5 to 9 units 4 - 10 to 19 units 5 - 20 to 49 units 6 - 50+ units
POPDEC2010_ACSX	2010 Census Population (Taken from 2013 MAF)	Number

Table 6. Case Prioritization Model Parameters

Parameter	Source	Description
init_prop	Fixed Frame File	A case's initial propensity to respond (predicted by 2010 decennial contact history information in the Day 0 Model)
group	Fixed Frame File	Study treatment
mu	Derived from Fixed Frame File	Whether or not the sample unit is in a multi-unit structure
how_contacted	Contact History Table	Mode of each contact attempt (telephone or in-person)
totalcount	Derived from Contact History Table	Total number of contact attempts already made on the sample unit
prior_contact	Derived from Contact History Table	Whether or not contact was made with a household member during the current or any previous contact attempts
prior_reluct	Derived from Contact History Table	Whether or not a potential respondent expressed reluctance during the current or any previous contact attempts
peak	Derived from Contact History Table	If the contact was performed during "peak" hours, including the weekend or after 6:00 p.m. on a weekday
hu_more_than_one	Fixed Frame File	Whether or not there is more than one person in the housing unit
hu_all_lt30	Fixed Frame File	Whether or not all sample unit members are less than 30 years-old
hu_all_gt70	Fixed Frame File	Whether or not all sample unit members are 70+ years-old
hu_kids_lt5	Fixed Frame File	Whether or not there are children under 5 years-old in the house

Before running this model, the program checked for at least one completed CAPI case. If one CAPI case had not yet been completed, all cases were set to the same propensity to respond. After running the model, all contact attempts were scored, including proxy contact attempts. Contact attempts were then ordered from newest to oldest, and the most recent contact attempt was selected. Cases that had not been worked were assigned their initial propensity from the Day 0 Model. All cases were then ordered from highest to lowest propensity for each CAPI interviewer, and a random number was attached to each case. The highest seven propensity cases were assigned "high priority" for each interviewer. In the case of a propensity score tie or

multiple ties, the highest random numbers were chosen until each interviewer had up to seven high priority cases.

3.4.2 Reduced Contact Attempt Code

The program also ran code and business rules that indicated to CAPI interviewers which cases were "Proxy if no Interview" and removed cases from their workloads for which a proxy interview was not obtained. A series of "counters" in the system summed the number of contact attempts for each case and sent a "Proxy if no Interview" instruction on the morning after an interviewer reach their second-to-last contact attempt. Thus, the "Proxy if no Interview" instruction was sent on an interviewer's final contact attempt. If an interviewer had seen a "Proxy if no Interview" icon and then performed a personal visit on a case but was not able to close it, the case was removed from his or her laptop. So as not to penalize the interviewer, cases were only stopped due to maximum attempts if the interviewer had seen a "Proxy if not Interview" icon already. (The icon may not have shown up due to a transmission error or working the cases several times in one day.) Cases were not stopped that had an interview set in the future, unless the appointment was broken.

3.4.3 Execution of Business Rules

The program generated a file each morning that communicated all necessary business rules. Each case in ROSCO (i.e., each CAPI case in the test) was assigned three pieces of information: whether or not the case was a high priority, whether or not the case was a "Proxy if no Interview," and whether or not the case should be stopped due to maximum attempts. UTS placed the file each morning onto the Census Bureau's Master Control System (MCS). MCS then processed the file and provided the data to ROSCO. ROSCO processed the data and created files for interviewers that were placed on the CAPI Telecommunications server for the interviewers to pick up on transmission. Interviewers were instructed to make a transmission every day before they started working to pick up these instructions.

3.4.4 Systems Testing

Before going to the field, the 2013 Census Test performed two systems tests, both of which tested the above program. Because the 2013 Census Test business rules relied on brand new ROSCO variables to execute correctly, the code could not be run on data, real or synthetic, before the first systems test. Rather, the first systems test was used to identify oversights in the code and generate data for further code development. The second systems test used these data to debug and refine the program.

3.5 Telephone Methodology

In the 2013 Census Test, telephone interviewing was implemented in two ways. In the adaptive treatments (3 and 4), phone calls were managed in a centralized CATI facility. In the fixed treatments (1 and 2), the telephone calls were managed by individual interviewers. These

different mode sequences represent two possible ways in which telephone and in-person data collection could be organized for the 2020 Census. The outcomes achieved in the adaptive and fixed treatments are the result of the combination of different telephone and in-person approaches.

The TCC was used to conduct the CATI operation for the adaptive group cases with telephone numbers. Cases were sent to a centralized calling queue managed by WebCATI. WebCATI uses previous call data and pre-determined parameters to order cases in the calling queue. Cases can go to any interviewer who is working when that case comes up in the queue. Cases not completed in CATI were then assigned to field representatives for CAPI interviewing.

Cases with phone numbers in the fixed design were assigned to CAPI interviewers. The phone numbers were displayed on their laptops in case management fields. For cases with telephone numbers, CAPI interviewers were instructed to attempt each phone number twice prior to attempting a personal visit.

For both treatments, cases might have had up to three telephone numbers. The system used for CAPI case management could not be modified to allow for more than three telephone numbers, so, in order to compare treatments, the upper limit was set at three phone numbers for both treatments.

3.5.1 Phone Number Selection

The phone numbers used in the 2013 Census Test came from the Center for Administrative Records Research and Applications (CARRA). Between 2010 and 2013, the Census Bureau purchased twenty-four commercial databases from seven vendors[5] containing name, address, and telephone numbers amongst other socio-demographic, geo-demographic and socio-economic variables for households in the United States. The address data on the commercial files were matched to the Census Bureau's Master Address File (MAF) using probabilistic record linkage software, MAFMatch, developed by CARRA. Linked to the MAF, the address and telephone number data in the commercial data comprise the Supplemental Contact Frame (CF). There are no Federal, Census Bureau, or survey data on the CF.

CF data were developed by linking commercial databases with address and telephone numbers to the MAF through MAFMatch. The MAF's unique identifier is the MAFID, a distinct address unit. During MAFMatch, a commercial data address was or was not appended to a MAFID. There is no record on the CF without a MAFID.

The CF is organized at the MAFID-telephone number unit of analysis, which means no row of data contains more than one telephone number or MAFID and no row of data is without both MAFID and ten-digit telephone number. There are no duplicate MAFID-telephone number combinations, but telephone numbers may be associated with one or more MAFIDs, and

[5] The seven third party data vendors were Experian, InforUSA, Lexis Nexis, Melissa, Targus, Telematch, and VSGI.

MAFIDs may have one or more telephone numbers. Each MAFID-telephone number combination contains flags that denote where the information came from.

CARRA's Telephone Number Type Assignment Process (TeNTAP), developed based on North American Numbering Plan Administration (NANPA) and Federal Communications Commission (FCC) data, classifies a phone number as a landline or a cellphone. Both landline and cellphone numbers were eligible for contact in the 2013 Census Test. After obtaining a list of phone numbers associated with sampled addresses, Census used PhoneTree software to remove nonworking landline telephone numbers. PhoneTree software, using auto-dialer technology, appended a "call status" to each landline phone number it was specified to call. Census associated three call statuses[6] as out of service. Landline telephone numbers determined likely to be out of service were removed from the phone numbers associated with the sample cases.

CARRA ranked the remaining landline and cellphone numbers for each case. Where there were multiple phone numbers associated with a sample address, we used logistic regression models to predict the types of phone records that were more likely to lead to contact with the sample address. These models used the results of the 2013 National Census Contact Test, where the objective was to confirm the accuracy of phone-address links, as a guide to predicting contacts with sample addresses (Sheppard, 2014). Models used three dummy independent variables: vendor source, year of the telephone number data, and frequency of the telephone number occurrence across multiple sources. The top three ranked phone numbers for each sample address were appended to the sample cases.

Table 7 shows the number of cases with at least one phone number by group prior to employing administrative records to remove cases from the workload.

Table 7. Total Number of Sample Units with at Least One Telephone Number by Group before Administrative Records (AR) Removals

	Fixed	Adaptive
AR removal treatment	360	408
No AR removal treatment	351	380
Total	**711**	**788**

Table 8 shows the number of cases with phone numbers in each treatment after cases were removed with administrative records (the top row of the table). 609 cases in the adaptive group were provided to the TCC for centralized CATI calling. 552 cases with telephone numbers were provided to CAPI interviewers in the fixed group.

[6] The call statuses used to determine likely out of service numbers were Call Failed!, Telco Message, and No Connect.

Table 8. Number of Sample Units with Telephone Numbers by Group after Administrative Records Removals

	Fixed	Adaptive
AR removal treatment	201	229
No AR removal treatment	351	380
Total	**552**	**609**

3.6 Resources

To conserve resources, the Census Bureau used existing staff and office infrastructure for the 2013 Census Test.

The Census Bureau recruited CAPI interviewers who had recently worked on the American Housing Survey. Two field supervisors from the Philadelphia Regional Office were chosen to manage these interviewers, one for the group working in the adaptive case management group and the other for the fixed group. Interviewers were randomly assigned to these groups. To obtain a balanced driving distance between the two groups, we conducted a driving distance analysis. Using Google map driving directions, we mapped the geocoded home address of each interviewer to each block pairing and calculated average distance traveled. We then moved several interviewers to the other group to balance average distances between block groups.

The Census Bureau used existing American Community Survey (ACS) systems for the 2013 Census Test, modifying where necessary. The test used the ACS case management system for displaying and recording case information. The test used a modified 2013 ACS production questionnaire, from which all person-level questions except relationship, sex, age, date of birth, Hispanic origin, and race were removed. The instrument contained only two household-level questions, tenure and status of temporarily occupied units. The test also used the Census Bureau's Contact History Instrument (CHI) to track contact attempts, strategies, and reluctance reasons. The case management system, production instrument, and contact history instrument were made available to CAPI interviewers on laptops.

The test relied on a modified ACS CAPI interviewer training because it used the ACS case management system, modified production instrument, and contact history instrument. This interviewer training recapitulated information provided to ACS interviewers but emphasized the importance of interviewer behavior in following contact procedures prescribed for the test and in recording contact history information.

In addition to using existing ACS infrastructure, the test also utilized other Census Bureau resources. Paradata were obtained on a daily basis using the Unified Tracking System (UTS). UTS is a centralized data repository for paradata and cost and progress data for several Census Bureau surveys. Its purpose is to track survey operations including case workloads and different modes. In addition to obtaining paradata, UTS was used as the production environment in which

propensity scores were modeled and business rules were executed. The 2013 Census Test also utilized existing CAPI and CATI cost recording mechanisms.

The 2013 Census Test developed its own CAPI interviewer performance reports. These daily reports tracked if interviewers followed prescribed contact procedures. Both adaptive and fixed groups were monitored on how frequently they transmitted their data, when cases were closed out after the maximum attempts were made, and how faithfully interviewers recorded contact history information. Interviewers assigned to the adaptive case management group were also monitored on contact attempts for high priority cases. Interviewers in the fixed group were monitored for their use of telephone before in-person contact attempts.

3.7 Assumptions

The 2013 Census Test key assumptions are listed below:

- The total sample size will be 2,000 cases, 500 in each treatment.
- The NRFU field period will last six weeks.
- The ACS survey instrument and systems will be modified for use for this test.
- The adaptive case management group (Treatments 3 and 4) will use a centralized outbound CATI system before the field enumeration.
- CARRA will provide up to three phone numbers for each housing unit.
- Transferring incomplete cases from CATI to CAPI interviewers will occur in a staged fashion, so interviewers receive groups of cases at a time.
- CAPI interviewers will pick up data at the beginning of each shift and transmit data at the end of each shift.
- Researchers will access and run models on data and paradata each night throughout the course of the field period.
- Based on business rules and the results of these models, each night the control system will apply a priority score and other instructions to each case in the field in the adaptive treatments using models.
- The adaptive group and the fixed group will each have one “Crew Leader District” (CLD). These CLDs will be comparable.
- Paradata and payroll data will provide the necessary information to calculate cost per case and data quality.

3.8 Research Questions

The 2013 Census Test piloted novel uses of administrative records and contact strategies to reduce NRFU workload and increase NRFU efficiency (i.e., interviewer productivity). It primarily tested the operational efficiency of innovative procedures applied for the first time in the field. As such, the data analysis focuses on how well systems and people worked to affect the study components. Much of this analysis is descriptive.

To examine the use of administrative records to identify vacant housing units and enumerate occupied housing units, we asked the following questions:

- Can we use administrative records and postal information to identify occupied and vacant housing units and enumerate occupied housing units?
- Alternatively, can we use records to determine the number of contacts (i.e., level of effort) for occupied sample units?
- How well do administrative records and postal information identify occupied and vacant housing units?
- How well do alternative sources identify vacant housing units and addresses that are not housing units?
- How well do administrative records do in identifying population counts for occupied housing units?

To examine the use of a reduced number of contact attempts, we asked the following questions:

- How many cases were completed in CAPI, and how many cases were incomplete?
- Of incomplete cases, how many could have their status determined and, if occupied, be enumerated using administrative records?
- What do we know about incomplete cases with no administrative records information?
- How many personal visit attempts were there per treatment?
- What percentage of occupied housing units were completed using proxy respondents?

To examine the use of an adaptive design contact strategy to prioritize cases, we asked the following questions:

- Can a program generate daily propensity scores and determined when to do proxy interview attempts?
- Can systems deliver daily high priority case designations to in-person interviewers?
- Will CAPI interviewers comply in manually transmitting data, recording contact attempts, performing proxy interviews, and working high priority cases as trained and supervised?
- What are the contact and cooperation rates between regular and high priority cases?

To examine the use of matched telephone numbers, we asked the following questions:

- Can we use centralized CATI and dispersed CAPI phone calls to enumerate sample units?
- How many cases with matched telephone numbers were completed in CATI and in CAPI by interviewers making telephone calls before personal visit attempts?
- What were the final outcome codes for cases completed by telephone?
- What were the final CATI outcome codes for cases attempted in CATI?

- What were the outcomes of CAPI phone attempts?

A secondary goal of the test was to examine cost, effort, and data quality differences between treatments. This study compared these dependent variables between treatments that used and did not use administrative records to reduce the NRFU workload and between treatments that used an adaptive versus a fixed contact strategy. Specific research questions were:

- How does the overall cost of CAPI and CATI interviewers compare between treatments?
- How does the average cases per hour compare between treatments?
- How does the average attempts per hour compare between treatments?
- How do completion rates compare between treatments?
- How do item nonresponse rates compare between treatments?
- How do percent proxy responses compare between treatments?

3.9 OMB Clearance

On August 15, 2013, OMB approved the 2013 Census Test under OMB Control Number 0607-0975.

3.10 Schedule

Table 9. Milestone Schedule Activities

Activity	Date
2020 Field Test Plan – 2013 Census Test finalized	7/24/2013
CATI interviewers trained	9/30/2013
Mail materials finalized	10/17/2013[†]
Instrument completed	10/17/2013
Initial letter mailed	10/18/2013
CAPI interviewers trained	10/31/2013-11/1/2013
Data collection began	11/2/2013
Data collection ended	12/5/2013
Finalized analyses	3/31/2014
Prepared draft 2013 Census Test Assessment	4/14/2014
Reviewed 2013 Census Test Assessment report and recommendations	5/15/2014
Updated 2013 Census Test Assessment report and recommendations	5/29/2014
Reconciled 2013 Census Test Assessment updates with knowledge management	5/29/2014
Issued final 2013 Census Test Assessment	5/30/2014

[†]: Mail materials were finalized after the study plan due to a Federal government shutdown that occurred from October 1, 2013 – October 16, 2013. This shutdown affected the 2013 Census Test plan and schedule, as the initial letter was prepared to be mailed on October 1, 2013.

4. Limitations

There were limitations related to the overall study design, use of administrative records, adaptive design intervention, and telephone contact mode.

4.1 Study Design

- The test was limited to one urban data collection region, so results are not generalizable to the entire country.
- The sample size (both housing units and interviewers) limited the ability to perform statistical tests across treatments.
- The test did not contain a self-response component. Instead, housing units that did not mail back a self-response form during the 2010 decennial Census were eligible for inclusion. Thus, the NRFU population was simulated using information from the 2010 Census.

4.2 Use of Administrative Records to Reduce Workload

- A limited amount of time (three months) was available to identify occupied and vacant housing units via administrative records in this test.
- The test had only one mailing – an advance letter that was sent only fifteen days before the first day of CAPI data collection. Thus, there was limited UAA information available to identify vacant housing units.

4.3 Adaptive Case Management

- Changes to case management were limited due to the reuse of existing ACS systems.
- The adaptive design model was limited by the small number of sample cases and maximum number of contact attempts.
- CAPI interviewers used existing laptops rather than the handheld devices planned for use in other 2020 field tests.
- The test was conducted in late 2013, but administrative record data was available only from 2011 and 2012.

4.4 Telephone Contact Mode

- The system used for CAPI case management could not be modified to allow more than three telephone numbers, so all treatments were limited to three phone numbers.
- The system used for CAPI case management could not be modified to automatically record CAPI phone attempts, so we relied on in-person interviewers to indicate in the CHI when phone calls were made. Recording telephone contacts was not accurate in all cases.
- The system used for CATI case management could not be modified to rotate automatically through the multiple phone numbers for each subsequent call, potentially lessening the ability to learn if the use of all available phone numbers would increase the possibility of completing interviews by phone.
- The test did not contain a self-response component, so the analysis could not measure the potential benefit of talking to respondents or leaving phone messages that lead to late mail returns.

5. Results

5.1 Using Administrative Records to Reduce NRFU Workload

This section reviews the use of administrative records to identify the occupancy status of sample units and the use of records to enumerate occupied units.

5.1.1 Using Administrative Record Information to Identify Vacant Units

We are interested in how often we correctly assigned a vacant status as well as how often we incorrectly assigned a case as vacant. To understand the validity of our assignment method,

Table 10 shows the final status results for Treatments 2 and 4. We focus on Treatments 2 and 4 because interviews in the field were completed for those treatments. Hence, we could compare the status assigned by our methods described in Section 3.3 against the field interview status.

Table 10. Field Status of Cases Assigned a Vacant Flag

Treatment	Units Assigned "Vacant" Status	Units found to be Occupied	Units found to be Vacant	Units found not to be Housing Units	Units with Incomplete Status
2	31	2	19	7	3
4	71	2	49	10	10

We assigned a vacant status correctly for 67.9% (19/28) and 80.3% (49/61) of the cases in Treatments 2 and 4. Since we flagged these units as vacant, and they were determined to be so in the field, these units can be thought of as true positives. We incorrectly assigned a vacant status to 7.1% (2/28) and 3.3% (2/61) of units in Treatments 2 and 4. These errors can be thought of as false positives (i.e., the cases for which we assigned a vacant status but were occupied). It is interesting to note in retrospect that three of the four units had persons in the unit according to the three administrative records sources used. However, the PIKs were not the same across multiple years within the same administrative records source, so they did not receive an initial occupied flag.

Now, let us look at the false negative rate (i.e., the cases to which we did not assign a vacant flag but were found to be vacant in the field). To start, Table 11 displays the 31 and 71 cases that we flagged as vacant in Treatments 2 and 4. Subtracting from the total, there are 479 cases we did not flag as vacant on Treatment 2 and 457 in Treatment 4.

Table 11. Vacant, Non-Vacant Assignments

Treatment	Overall Total	Vacant Flag Assigned	Did Not Flag As Vacant
2	510	31	479
4	528	71	457

We would like our methods to identify as many vacant units as possible prior to doing NRFU interviews. However, there were some vacant cases that our procedures missed and were

determined to be vacant after the NRFU interview. Table 12 shows the final status results for Treatments 2 and 4 for cases we did not assign as vacant.

Table 12. Field Status of Cases not Assigned Vacant Status with Records

Treatment	Units Not Assigned "Vacant" Status	Units found to be Occupied	Units found to be Vacant	Units found not to be Housing Units	Units with Incomplete Status
2	479	279	55	26	119
4	457	243	69	44	101

In Treatments 2 and 4 respectively, 55 and 69 of the cases were vacant. Further analysis suggests some reasons why we did not assign vacant status to these units before they reached the field. Some of the reasons include:

- There was no UAA information from the Postal Service.
- The UAA had a "reason code" for delivery failure other than "vacant."
- Other reasons:
    - The UAA vacant reason code designation was received after the case was sent to the field. (Cases were assigned to the field on October 29, 2013.)
    - Cases were assigned an UAA vacant reason code, but other administrative records sources indicated that it was occupied.

Table 13 displays the reasons why a vacant case was not assigned as vacant. Of the 55 cases in Treatment 2 that were found vacant during interviewing, 45 had no UAA information. Similarly, 50 of the 69 in Treatment 4 had no UAA information.

Table 13. Reasons for Failure to Assign Vacant Status

Treatment	Units	No UAA Information	UAA Reason Code other than Vacant	Other Reasons
2	55	45	9	1
4	69	50	6	13

Based on these findings, expanding the range of UAA reason codes used to assign vacant status seems worth further investigation.

5.1.2 Identifying Vacant Units by Delivery Point Validation – Vacant Status

For this test, we combined the presence of an UAA – Vacant flag and the absence of contradictory administrative record data to remove vacant addresses from the workload. While the technique was not used in this test, before sending a mailing it is also possible to perform a

delivery point validation (DPV) of an address.[7] We looked retrospectively at whether DPV – Vacant status would have been helpful in identifying vacant units. Table 14 provides statuses of cases assigned as vacant by DPV – Vacant status.

Table 14. Field Status of Cases by Delivery Point Validation – Vacant Classification

Treatment	DPV – Vacant Status	Sample Units	Units found to be Occupied	Units found to be Vacant	Units Found not to be Housing Units	Units with Incomplete Status
2	Yes	42	8	18	9	7
	No	468	273	56	24	115
	Treat. Total	**510**	**281**	**74**	**33**	**122**
4	Yes	92	7	47	29	9
	No	436	238	71	25	102
	Treat. Total	**528**	**245**	**118**	**54**	**111**

When a DPV – Vacant status was indicated, units were found to be vacant more often. For Treatment 2, 51.4% (18/35) of DPV - Vacant units were found to be vacant as opposed to 15.9% (56/353) of units without a DPV – Vacant code. For Treatment 4, 56.6% (47/83) of DPV - Vacant units were found to be vacant as opposed to 21.3% (71/334) of units without a DPV – Vacant code. This results show that the DPV – Vacant code adds additional justification for a vacant classification.

5.1.3 Using Administrative Record Information to Identify Occupied Units

We are interested in how often we correctly and incorrectly assigned occupied statuses to cases. To understand the validity of our assignment method, Table 15 shows the final status results for Treatments 2 and 4.

Table 15. Field Status of Cases Assigned an Occupied Flag by Treatment

Treatment	Units Assigned "Occupied" Status	Units Found to be Occupied	Units Found to be Unoccupied (Vacant or Not a Housing Unit)	Units with Incomplete Status
2	150	94	17	39
4	152	91	13	48

[7] DPV is a service available from the United States Postal Service to mailers such as NPC. Its purpose is to identify whether or not a piece of mail can be delivered to a specific address.

We assigned occupied status correctly 84.7% (94/111) and 87.5% (91/104) of the time in Treatments 2 and 4. Since we flagged these units as occupied, and they were determined to be so in the field, these units can be thought of as true positives.

5.1.4 Generating Population Count from Administrative Record Sources

For units found to be occupied during the interview, we compare the administrative records count to the interview count. For this research, we take a conservative approach by restricting the administrative records population count to come only from units that were flagged as an administrative records' source. According to the rules we employed, all persons had to be recorded as living in the unit for at least two years. By doing so, we exclude persons associated with the unit who may be in an administrative records source only once. Table 16 shows two examples of how counts would be generated using administrative records sources.

Table 16. Two Examples Generating Household Population Counts with Administrative Records (AR)

Housing Unit	IRS 2011	IRS 2012	Medicare 2011	Medicare 2012	Targus 2011	Targus 2012	AR Source	AR Pop. Count
101	A,B	A	C	C	A,D	A,E	Medicare Only	1
102	F,G	F,G			H	H	IRS and Targus	3

For Housing Unit 101, Medicare is the only administrative records source with the same persons in the unit across both years. Hence, it is listed as the administrative records source. The administrative records population count is "1" because Person C is the only person listed on an administrative records source where all persons are the same across two years. Although Person A is on the IRS and Targus sources for two years, Person A is excluded because not all persons match in each of the two sources. For this initial test, we assigned an occupied flag only if all people in the household matched across the two years for any given administrative records source. Although Person A is on the IRS file over two years, the IRS data were not used to flag the unit as occupied. Hence, we did not include Person A as part of the administrative records household roster. (We have been researching other ways to confront this scenario going forward.) For IRS, Person B is in the 2011 version and not in the 2012 version. For Targus, Person D is in the 2011 version and not in the 2012 version, and Person E is not in the 2011 version but in the 2012 version. The lack of person agreement across both years for IRS and Targus disqualifies them from being listed as an administrative records source. Furthermore, Person A is not included in the count despite being in two administrative records sources across both years.

For housing unit 102, IRS and Targus list the same persons in the unit across both years. Even though the persons are not the same across the two administrative records sources, the administrative records population count is three because Persons F, G, and H are in an administrative records source across two years. In general, for units in which we had multiple administrative records sources, the administrative records population count was constructed from the union of the administrative records sources.

Table 17 compares the household size counts of occupied housing units true positives between the administrative records source (i.e., administrative records count) and as determined through the NRFU interview (i.e., NRFU count).

Table 17. Household Size Comparison between Administrative Records and NRFU Interview by Administrative Records (AR) Source

Has Excluded AR Persons	AR Count 2+ Less than NRFU Count	AR Count 1 Less than NRFU Count	AR Count Equal to NRFU Count	AR Count 1 More than NRFU Count	AR Count 2+ More than NRFU Count	Interview Occupied but NRFU Count not Determined	Total
Count	41	32	61	22	17	12	185
Percent	23.7%	18.5%	35.3%	12.7%	9.8%	-	100.0%

Table 17 shows 42.2% (73/173) of the housing units have a household size based on administrative records that is smaller than the NRFU interview. However, 58.9% (43/73) of these records have some administrative records persons that were excluded when building the administrative records count. The table also shows 66.5% (115/173) of the housing units have a household size within one of the NRFU interview. As noted in Section 4. Limitations, this test used administrative records source data from 2011 and 2012 while NRFU counts were collected in late 2013. Some of the differences between administrative records and NRFU counts might be corrected in future tests where the administrative records source data are more current with Census Day.

In Table 17, the distribution in the left columns is higher than the distribution in the right. This suggests that the administrative records count is usually less than the interview count. It is possible that the administrative records count could be less because of the conservative approach we took in generating the administrative records count as illustrated in Table 16. We continue to research whether excluded persons like those in Table 16 should be included in the administrative records count. For certain administrative records sources, it may be sufficient to be in the source for the most recent year to be part of the administrative records count.

Now, let us look at the false negative rate (i.e., the cases that we did not assign an occupied status to but were occupied in the field). To start, Table 18 displays the 150 and 152 cases that

we flagged as occupied in Treatments 2 and 4. Subtracting from the total, there are 360 cases we did not assign as occupied in Treatment 2 and 376 in Treatment 4.

Table 18. Occupied/Non-Occupied Assignments

Treatment	Units	Occupied Flag Assigned	Did Not Assign Occupied Flag
2	510	150	360
4	528	152	376

We would like our methodology to be able to identify as many occupied units as possible prior to doing NRFU interviews. However, there were some occupied cases that our methodology missed and were determined to be occupied after the NRFU interview. Table 19 shows the final status results for Treatments 2 and 4 for cases we did not assign as occupied.

Table 19. Field Status of Cases Not Assigned an Occupied Flag with Records

Treatment	Units Not Assigned "Occupied" Status	Units found to be Occupied	Units found to be Vacant	Units found not to be Housing Units	Units with Incomplete Status
2	360	187	62	28	83
4	376	154	108	51	63

In Treatments 2 and 4 respectively, 187 and 154 of the cases were occupied. Further analysis suggests some reasons why we did not assign these records as occupied before they reached the field. Some of the reasons include:

- There were no PIKs in any of the three administrative records sources.
- PIKs were in at least one of three administrative records sources, but the PIKs were not identical over the two years.
- PIKs were identical in at least one of three administrative records sources over the two years. However, unvalidated person records existed for that unit.

Table 20 shows the reasons why an occupied case was not assigned as such from administrative records sources.

Table 20. Reasons for Failure to Assign Occupied Status

Treatment	Units	No PIKs - Any Source	PIKs – 1+ Source, Not Same Over 2 Years	PIKs – Same 1+ Source, Unvalidated Person Records for Unit
2	187	71	96	20
4	154	41	100	13

The majority of the cases in both treatments are those for which we had PIKs that did not agree across the two years. This result suggests that we may want to re-evaluate the criteria applied to remove occupied cases from NRFU work based on administrative records sources.

5.2 Reducing Maximum Personal Visit Attempts to Increase NRFU Efficiency

The 2013 Census Test examined reducing the number of maximum personal visit attempts. In the 2010 decennial Census, enumerators were permitted up to six contact attempts for each case. In comparison, the 2013 Census Test used a maximum of three personal visit attempts in Treatments 1, 2, and 3. In Treatment 4, the presence or absence of administrative records suitable to enumerate the household determined the level of contact effort.

5.2.1 Contact Attempts

Cases had a varied number of personal visit attempts. In Treatments 1, 2, and 3, all cases were allowed up to three personal visits. For cases with record information available to enumerate the household in Treatment 4, in-person interviewers were permitted one personal visit attempt to enumerate the household. For cases without record information, up to three personal visit attempts were permitted.[8]

The case management system instructed CAPI interviewers when it was their final personal visit attempt. If an interviewer could not complete an interview during this attempt, they were trained to attempt a proxy interview for that housing unit. In addition to serving as respondents for occupied housing units, proxies verified vacant housing units and, where possible, addresses that were not housing units. If interviewers could not obtain proxy interviews on their final personal visit attempts, the cases were removed from their laptops at the time of the next transmission.

Table 21 shows the number of personal visits attempts by treatment for various types of cases. Complete cases include completed cases with data that were worked in CAPI as well as

[8] Cases may have received more than three personal visit attempts because an appointment was set on the final personal visit attempt or multiple personal visit attempts were done on the last day the case was in the field. Furthermore, in-person interviewers may not have transmitted correctly, causing interviewers to be able to attempt cases more than the maximum attempts allowed.

completed cases without data (Type As and Type Cs[9]). Incomplete cases were stopped due to "maximum attempts" for which no data were collected in the field, including cases that could be enumerated with administrative records. Treatments 2 and 4 show a larger total number of personal visit attempts than Treatments 1 and 3 because the workload was reduced with record information in the latter groups. The number of visits per case, however, was higher in Treatments 1 and 3, which means cases remaining after some were removed using administrative records information are more difficult.

CAPI interviewers were instructed to try to obtain interviews from proxy respondents after each case's final day in the field. Due to these proxy attempts, incomplete cases show higher average personal visit attempts than complete cases (3.31 versus 2.24 attempts).

Table 21. Personal Visit (PV) Contact Attempts by Treatment

Treat.	CAPI Cases Worked	Complete Cases	Incomplete Cases	PVs	PVs per Case	PVs per Complete Case	PVs per Incomplete Case
1	311	253	58	781	2.51	2.36	3.19
2	510	388	122	1248	2.45	2.21	3.20
3	313	259	54	853	2.73	2.40	4.30
4	508	397	111	1165	2.29	2.09	3.01
Total	**1642**	**1297**	**345**	**4047**	**2.46**	**2.24**	**3.31**

5.2.2 Final Contact Attempt and Proxy Respondents

Table 22 lists the respondent type for all CAPI cases identified as occupied and temporarily occupied.[10] In total, 31.5% of occupied or temporarily occupied housing units were closed with a proxy respondent, but 25% were closed with an unknowledgeable proxy respondent (i.e., a proxy respondent who only provided a population count on a housing unit).

Table 22. CAPI Respondent Type for Occupied & Temporarily Occupied Housing Units

Respondent Type	Number Cases	% Cases
Household	573	68.5
Proxy		
Knowledgeable	54	6.5
Not Knowledgeable	209	25.0
Total	**836**	**100.0**

[9] Type As include cases whose households could not be interviewed such as those with language problems, persistent absence, or refusals. Type Cs include cases that are not housing units such as demolished or condemned housing units and businesses for which a status is assigned but no interview is performed.

[10] The respondent type designation (i.e., proxy or household) was recorded in the survey instrument. However, the system did not save this variable correctly, so it was later obtained from trace files. Due to this reliance on trace files, respondent type may have been incorrectly recorded.

5.2.3 Incomplete Cases

If a CAPI interviewer made a final personal visit attempt but could not complete a case, it was removed from his or her laptop or stopped due to maximum attempts.[11] Of these stopped cases, some could be assigned a status with administrative records.

Table 23 shows complete and incomplete cases with administrative records flags by treatment. Treatments 1 and 3, all cases that could have been enumerated with administrative records had already been removed before going to field. In total, there were 345 incomplete cases. Of these, 245 contained no data, including any that could have been gleaned from administrative records.

Table 23. Cases with Administrative Record (AR) Flags by Treatment

Treatment	Total Cases	Case w/ AR Status Assigned	Cases Removed Before Data Collection	Incomplete Cases	Incomplete Cases w/ AR Status Assigned	Incomplete Cases w/ no AR Status Assigned
1	511	200	200	58	N/A	58
2	510	181	N/A	122	42	80
3	528	208	208	54	N/A	54
4	528	223	N/A	111	58	53
Total	**2077**	**812**	**408**	**345**	**100**	**245**

To learn more about these 245 incomplete cases without administrative record information, we analyzed available data sources including postal service information and contact history information.

Two of the 245 cases came back with errors from the DPV service performed by the National Processing Center (NPC) before mailout of the advance letter. Of these, one of the street names was not found in the DPV, and the other had a primary address that was not confirmed by the DPV. As shown in Table 24, another eleven of these 245 cases (4%) contained UAA mail classifications from the USPS, five of which indicated that the address might be vacant.

[11] Cases that had an appointment set in the future were not immediately removed.

Table 24. Incomplete Cases with No Administrative Records (AR) Information by UAA Reasons Code

UAA Reason Code	Incomplete – No AR Info.
Attempted, not known	3
Not deliverable as addressed	3
Vacant	5
Total	**11**

According to contact history records (Table 25), in-person interviewers reported making contact with 77 of the 245 cases (31%). See Table 25 for contact history information related to these incomplete cases. Interviewers reported that household members at 71 of theses cases (29%) reported some form of reluctance to participate in the interview, with 35 (14%) exhibiting reluctance on two or more visits.

Table 25. Incomplete Cases with No Administrative Records (AR) Information by CHI Indicators

CHI Indicators	Incomplete Units with No AR Info.
Only noncontacts with sample unit	168
At least one contact with sample unit member	77
Reluctance with sample unit	71
Language barrier/Other	6
Total	**245**

5.3 Implementing an Adaptive Design Case Management Strategy to Increase NRFU Efficiency

The 2013 Census Test examined the use of an adaptive design approach to manage CAPI case assignments. The following section discusses how well interviewers did at following instructions necessary to implement this adaptive approach and evaluates the case prioritization model.

5.3.1 CAPI Interviewer Compliance

The 2013 Census Test demanded that CAPI interviewers follow strict contact procedures. These included transmitting data to and from their laptops twice each day that they worked, filling out the contact history instrument, and performing proxy interviews after their final personal visit contact attempt. Interviewers in the adaptive case management group were also instructed and supervised to attempt high priority cases first.

5.3.1.1 Transmitting Data

It was necessary for CAPI interviewers to transmit before they began work each day and after they finished their work for the day. Transmissions push contact history information and outcome codes from interviewer laptops to the Regional Office Survey Control System (ROSCO). The paradata were used in the program that ran daily response propensity models and counted the number of personal visit attempts for each case. Transmissions also pull any updated interviewer instructions from ROSCO to the laptops as well as change the status of cases that have reached their maximum number of contact attempts on the laptops. Instructions are necessary to enact the correct contact procedures, which change daily.

Using laptops loaded with modified American Community Survey (ACS) instruments, all transmissions to download completed work and upload new work are manual. In-person interviewers were trained and monitored to transmit once before working, no earlier than 8:00 a.m., in order to receive their daily instructions. Interviewers were also trained to transmit once in the evening after they were done working, no later than 10:00 p.m., in order for the program to be executed with the most up-to-date contact history and outcome information.

Morning transmissions were judged compliant if an in-person interviewer transmitted after 6:00 a.m., before performing any contact attempts that day. Evening transmissions were judged compliant if an interviewer transmitted after performing any contact attempts but earlier than midnight. This is because transmissions allowed the correct data to be transmitted onto interviewer laptops if they happened between 6:00 a.m. and midnight each day.

Over 71 percent of transmissions were completed as trained (between 8:00 a.m. and 10:00 p.m.), and over 85 percent of transmission days were compliant, occurring between 6 a.m. and midnight. Table 26 shows the number and percent of compliant daily transmissions by each CAPI interviewer. Compliant transmission days (i.e., two transmissions between 6:00 a.m. and midnight, one before and one after all contact attempts that day) ranged between 33 and 100 percent by interviewer, with nine of 18 interviewers performing over 90 percent compliant transmission days and 12 interviewers performing over 75 percent compliant transmission days. Three interviewers had 60 percent or fewer compliant transmission days.

Table 26. Percent Compliant Daily Transmissions by CAPI Interviewer

Interviewer	% Compliant Daily (a.m. & p.m.) 6 a.m.-midnight	% Compliant Daily (a.m. & p.m.) 8 a.m.-10 p.m.
1	100.00	93.33
2	95.24	90.48
3	96.00	96.00
4	94.44	88.89
5	33.33	33.33
6	89.47	78.95
7	66.67	55.56
8	60.00	20.00
9	100.00	100.00
10	82.61	60.87
11	100.00	95.83
12	84.62	76.92
13	74.07	14.81
14	57.14	35.71
15	91.67	91.67
16	94.44	94.44
17	66.67	25.00
18	100.00	95.65
Total	**85.53**	**71.38**

In-person interviewers' transmission compliance generally ranged between 70 and 100 percent over the CAPI field period. Figure 1 shows the percent of interviewers who worked and transmitted correctly each day of the field period. On November 7, November 29, and December 3, compliance between 8:00 a.m. and 10:00 p.m. for working interviewers dropped below 50 percent. From November 27 onward, seven or fewer interviewers worked each day, so percentage compliant daily transmissions are based on very small numbers of working interviewers.

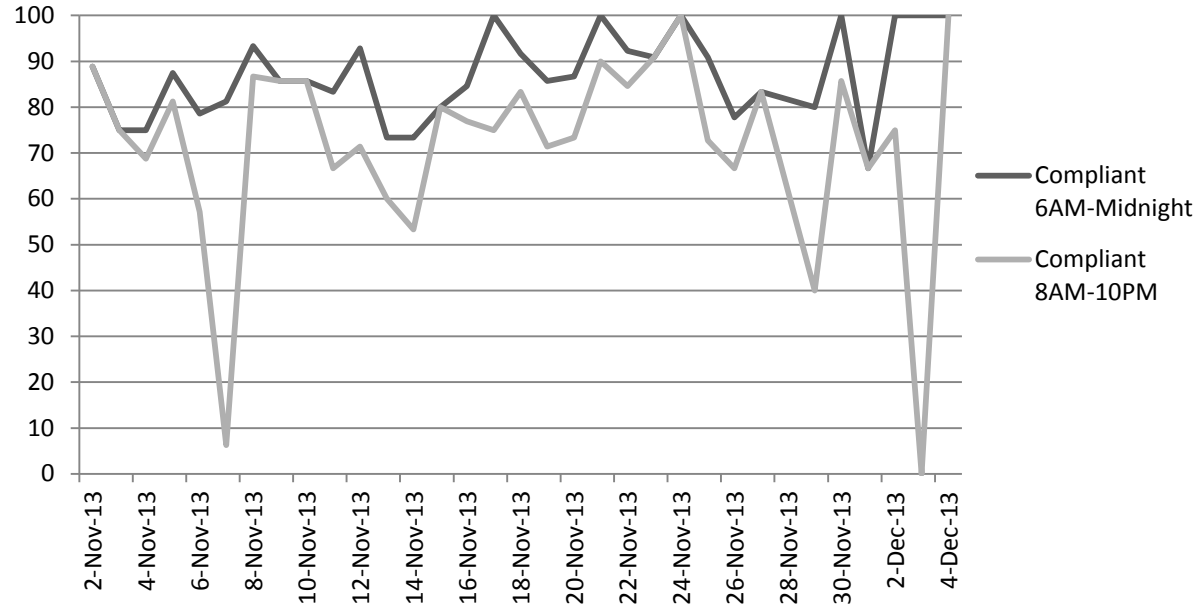

Figure 1. Percent Compliant Daily Transmissions Over CAPI Field Period

5.3.1.2 Recording Contact Attempts

The 2013 Census Test relied on CAPI interviewers to fill out the contact history instrument after each contact attempt. This information was used daily in the program. Specifically, contact records were used as input to the response propensity model and to determine the number of allowed personal visit attempts remaining.

While it is practically impossible to determine whether in-person interviewers filled out the CHI correctly, we can look at the timeliness of their records and the number of worked cases for which no contact records exist. Interviewers appeared to record contact attempts in a timely manner, reporting 97.18% of attempts to have been recorded during the time of the actual attempt.[12] In total, only six cases in the CAPI workload did not have a CHI record. One of these cases was completed as a vacant, and the other five were not housing units.

5.3.1.3 Performing Proxy Interviews

CAPI interviewers in all four treatments (in both adaptive and fixed groups) were trained and monitored when to get a proxy interview. Interviewers in both the adaptive case management and fixed groups were allowed up to three personal visit contact attempts, depending on the treatment and case. If an interviewer could not complete an interview on the last personal visit attempt, he or she was instructed to obtain a proxy interview by the end of that day. If no

[12] There is evidence that interviewers rounded self-reported times on the 2.82% of calls that they entered manually, with 64% recorded as occurring on minutes with a multiple of fifteen (0, 15, 30, and 45). Nearly a quarter of self-reported times (23.2%) had 0 as the minute.

interview was obtained, the case would be removed from the workload on the next day. These cases were referred to in training as "Proxy if no Interview."

The case management system communicated to CAPI interviewers when only one more contact attempt was allowed. An asterisk in a column next the control number indicated that a case was a "Proxy if no Interview." When selected, this case showed a "neighborhood" icon on the Assignment Tab in the Case Management System. (See Figure 2 for a screenshot.)

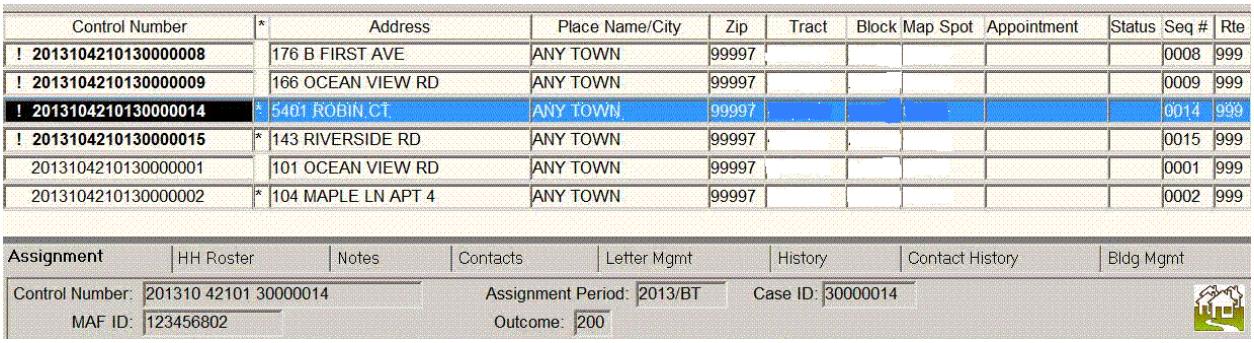

Figure 2. Screenshot of 2013 Census Test Case Management System

Analyzing if CAPI interviewers followed proxy interview instructions is difficult for the 2013 Census Test because contact history records do not differentiate noncontacts on proxy attempts from noncontacts with the sample household. We can only infer compliance from patterns of contact attempts when interviewers worked "Proxy if no Interview" cases including whether or not they performed multiple contact attempts to complete cases.

As shown in Table 27, over 62% of cases on their final personal visit attempt were completed.[13] (This analysis includes only cases that occurred on days with compliant transmissions, between 6:00 a.m. and midnight.) The percent completed on the first interview was 44.52%. As in-person interviewers should have made multiple contact attempts to complete cases they were not able to interview on the first attempt, we look at cases completed in multiple contact attempts. The percent completed with more than one contact attempt was 17.62%. However, 28.46% of cases on their final personal visit attempt were only attempted one time and were uncompleted. As the breakdown by interviewers show, some interviewers did not complete a majority of their final personal visit cases while others completed most of these cases. Furthermore, some interviewers had a majority of their cases not completed after only one personal visit attempt, indicating they did not put effort into closing such cases with a proxy interview. Based on supervisor and interviewer feedback, such interviewers may not have understood the "Proxy if no Interview" procedure.

[13] Cases not completed with appointments set in the future were not included in this analysis, as these cases would have received an additional contact attempt. Cases resulting in noninterviews (i.e. Type As and Type Cs) were not included in this analysis, as these cases may have gone into supervisory review and been eligible for an additional contact attempt or closed out by the supervisor. TQA completes were not included in this analysis, as these completions do not involve an interviewer's action.

Table 27. Percent Cases Completed versus Not on Final Personal Visit (PV) Attempt Day for CAPI Interviewers with Compliant Transmissions

Interviewer	Completed in One on Final PV Attempt Day	Completed in Multiple Attempts on Final PV Attempt Day	Not Completed after Multiple Attempts on Final PV Attempt Day	Not Completed after One Attempt on Final PV Attempt Day
1	12.82%	51.28%	12.82%	23.08%
2	40.45%	44.94%	10.11%	4.49%
3	93.33%	3.33%	0.00%	3.33%
4	16.90%	1.41%	7.04%	74.65%
5	38.98%	10.17%	3.39%	47.46%
6	62.50%	8.33%	0.00%	29.17%
7	84.21%	15.79%	0.00%	0.00%
8	64.81%	12.96%	16.67%	5.56%
9	77.78%	7.41%	0.00%	14.81%
10	29.36%	34.86%	19.27%	16.51%
11	40.00%	0.00%	11.43%	48.57%
12	44.44%	5.56%	5.56%	44.44%
13	20.93%	16.28%	6.98%	55.81%
14	25.81%	19.35%	41.94%	12.90%
15	43.64%	1.82%	0.00%	54.55%
16	0.00%	0.00%	0.00%	100.00%
17	100.00%	0.00%	0.00%	0.00%
Total	**44.52%**	**17.62%**	**9.40%**	**28.46%**

Note: There are only 17 CAPI interviewers listed because one did not have any final personal visit attempt days. Note: Cases with appointments set in the future and cases with outcome codes indicating Type As, Type Cs, or TQA completes were not included in this analysis.

Figure 3 illustrates how well CAPI interviewers followed this procedure over the field period. The percentage of cases completed on their final personal visit attempt appears to increase slightly over the field period. However, there were a very small number of such cases worked each day in the field, so trends are not apparent.

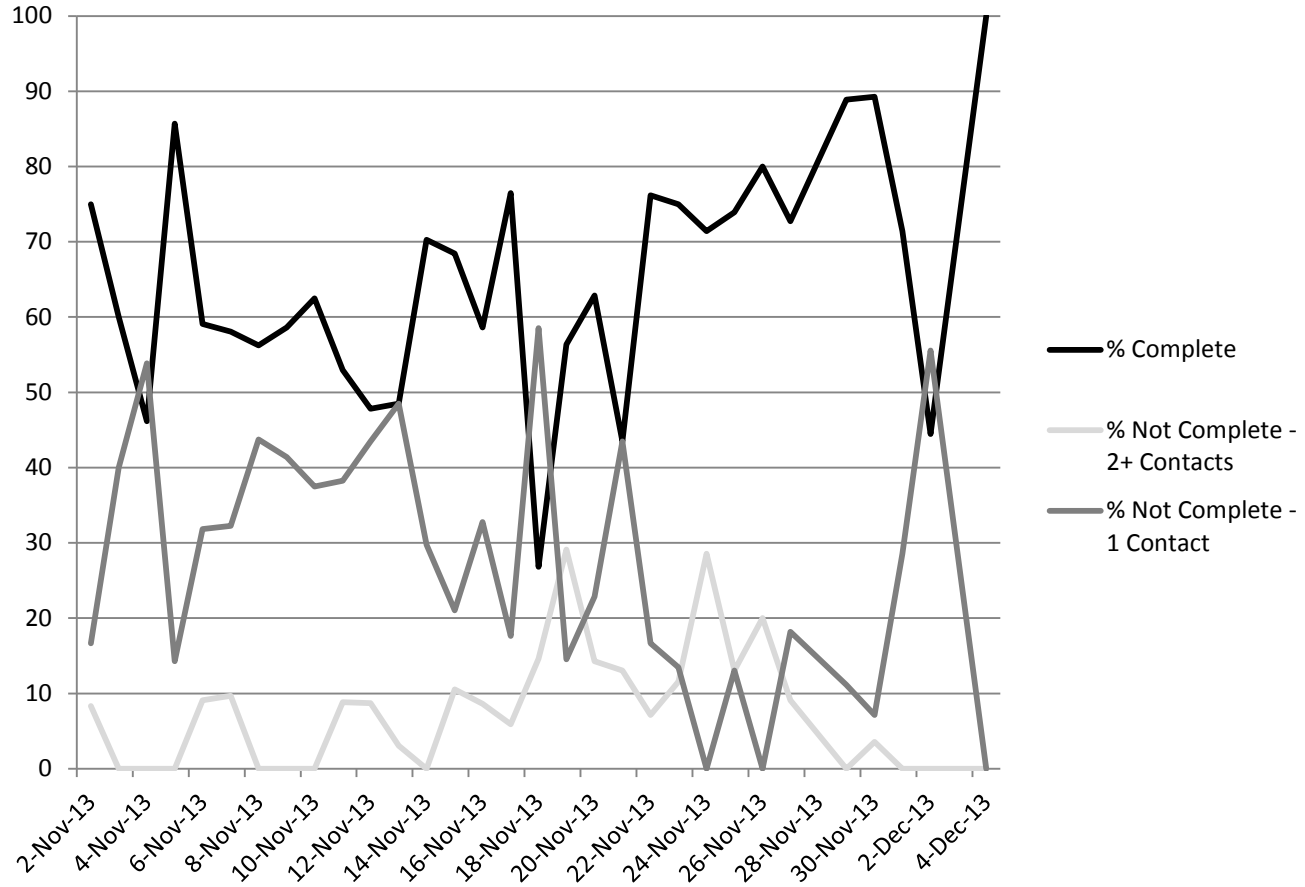

Figure 3. Outcomes of Final Personal Visit Attempt Cases Made by Interviewers with Compliant Transmissions Over Field Period

5.3.1.4 Working High Priority Cases

CAPI interviewers in the adaptive case management group (Treatments 3 and 4) received seven cases each day designated as "high priority."[14] These high priority cases corresponded to the seven highest propensity scores for cases not yet completed. Each high priority case was bolded on an interviewer's case management system, preceded by an exclamation point, and high priority cases were sorted to the top. (See Figure 2 for a screenshot.)

In-person interviewers in the adaptive case management group were trained to attempt all seven high priority cases on the days that they worked. They were allowed to visit other cases nearby and instructed to keep appointments. As shown in Table 28, interviewers worked all high priority cases (and may have worked other cases) on 45% of days with compliant transmissions, defined as between 6:00 a.m. and midnight. On nearly 7% of days with compliant transmissions, interviewers did not attempt all high priority cases but also did not attempt other cases, which may indicate that interviewers ran out of time to attempt all their high priority cases but were still following instructions to attempt high priority cases first. Interviewers did not attempt all high

[14] CAPI interviewers may have received more or fewer than seven cases due to cases being reassigned to other interviewers, not transmitting, or other anomalies.

priority cases and worked regular cases on nearly 48 percent of days when they transmitted correctly. Two interviewers worked all their high priority cases over 65% of the days they worked and were compliant with transmissions, while two interviewers attempted regular cases even though they did not attempt all their high priority cases over 60% of the days they worked and were compliant with transmissions.

Table 28. Percent High Priority Cases Worked Correctly by CAPI Interviewers with Compliant Transmissions (Treatments 3 and 4)

Interview.	% Days Attempted all High Priority Cases	% Days Did Not Attempt all High Priority Cases but Attempted No Reg. Cases	% Days Did Not Attempt all High Priority Cases and Attempted Reg. Cases
1	65.00	0.00	35.00
2	35.29	17.65	47.06
3	33.33	11.11	55.56
4	66.67	6.67	26.67
5	29.17	4.17	66.67
6	54.55	0.00	45.45
7	45.45	9.09	45.45
8	25.00	12.50	62.50
Total	**45.22**	**6.96**	**47.83**

5.3.2 Case Prioritization Model Evaluation

To test propensity score predictions made by the Case Prioritization Model, we compared contact and completion rates between high priority and regular (not high priority) cases in Treatments 3 and 4. Analysis included contact attempts made on November 4 and later, as November 4 was the first day all CAPI interviewers received cases prioritized by propensities from the Case Prioritization Model. To ensure that interviewers saw the correct high priority designation for each case, analysis was limited to days interviewers made compliant transmissions. Only personal visit contact attempts were included. Personal visit attempts with appointments set in the future were excluded from the analysis. Personal visit attempts resulting in a completed CAPI interview with a proxy respondent were also excluded, as these were omitted from the Case Prioritization Model.

A CAPI personal visit attempt was considered complete if the outcome of the attempt was occupied, temporarily occupied, vacant, a sufficient partial, or a telephone questionnaire assistance (TQA) complete. A personal visit attempt was considered a contact if the interviewer recorded contact with a sample unit or non-sample unit member (i.e., proxy).

Table 29 shows contact and cooperation rates for high priority and regular cases in Treatments 3 and 4. While contact rates are similar between regular and high priority cases in both treatments,

completion rates are lower for high priority cases compared to regular (not high priority) cases in both Treatments 3 and 4.

Table 29. Contact and Completion Rates between Regular and High Priority Cases

Treatment	Priority	Number	Contact Rate	Completion Rate
3	Reg.	316	0.259	0.206
	High	242	0.240	0.149
4	Reg.	388	0.224	0.209
	High	337	0.226	0.160

We examined several potential reasons why personal visit attempts to high priority cases had a lower completion rate compared to regular cases during the test. First, the Day 0 Model may not have estimated accurate case propensities. To examine this possibility, we reran the analysis excluding the first contact attempt on each case. Results showed a similar pattern of higher completion rates for regular cases compared to high priority cases.

Second, previous analysis (see Section 5.3.1.4, Working High Priority Cases) showed that adaptive interviewers attempted all high priority cases (plus other cases) fewer than 50 percent of the days they worked and successfully transmitted. If interviewers worked only one or two high priority cases each day, they may inadvertently have chosen high priority cases that were more difficult to interview. Table 30 shows the number of cases attempted each day by priority type (high priority or regular). Interviewers worked five or more high priority cases nearly half of the days they worked. It is therefore unlikely that interviewers consistently chose CAPI cases that were more difficult to complete.

Table 30. Number of Cases Worked per Day by Priority (Adaptive Only)

Number Cases Attempted by Interviewer	Priority	Percent Days
Less than 5	Reg.	54.17
	High	53.33
5 or more	Reg.	45.83
	High	46.67

Third, previous analysis (see Section 5.3.1.4, Working High Priority Cases) showed adaptive interviewers attempted regular cases but did not attempt all high priority cases nearly 48% of the days they worked. Interviewers may have chosen to interview regular cases when they were easy to complete. For example, interviewers may have performed interviews when they saw housing unit members at home and verified nearby vacant housing units when they had a willing proxy

respondent. Further analysis should examine why interviewers chose to attempt regular cases in an effort to explain observed completion rates.

5.4 Employing Telephone Calls to Increase NRFU Efficiency

CATI interviewing in the adaptive group was started on November 2, 2014 with an initial workload of 609 cases. Cases were scheduled to be called for up to two weeks with a maximum of 10 calls per case. CATI cases closed due to reaching the maximum of 10 call attempts, being refused twice, or having all numbers ruled ineligible were released in batches for field interviewing. These releases occurred on November 6 (202 cases), November 8 (159 cases), and November 13 (149 cases). CATI operations stopped on November 15, and any cases that were not complete were sent to field on November 16.

Phone numbers were available for 552 of the cases initially released to individual field interviewers in the fixed group on November 2. Interviewers were trained to attempt cases with phone numbers by telephone prior to making any personal visit attempts. There were instructed to try each phone number twice at different times of the day, prior to attempting personal visit interviews. If they were unsuccessful in obtaining a telephone interview after trying each available phone number twice, they were then instructed to attempt a personal visit attempt for that case.

Overall, interviewers completed 41 cases with initial telephone contacts: 27 by the Tucson CATI operation (adaptive case management group) and 14 by individual interviewers (fixed case management group) before any personal visit attempts were made to that case.

Table 31. Proportion of Completed Telephone Interviews by Treatment (Phone Interview Completed Before any Personal Visit Attempts)

	Fixed	Adaptive
AR removal	0.5%	3.1%
No AR removal	3.7%	5.3%
Total	**2.5%**	**4.4%**

Table 32 shows the final status for all cases sent to the field with a telephone number. For the adaptive case management group, this includes all cases that were initially sent to centralized CATI. For the fixed treatment, this includes all cases with telephone numbers that were sent to CAPI interviewers. Of all cases with phone numbers that were sent to centralized CATI, 18.9% had no data collected[15], 19.4% had a final outcome of vacant or temporarily occupied, and 5.8% were not housing units. For the cases sent to CAPI with phone numbers, 19.8% were stopped with no data collected, 12.7% had a final outcome of vacant, and 4.4% were not housing units.

[15] This category includes incompletes that were stopped due to maximum contact attempts as well as non-interviews, "Type As" whose households could not be interviewed such as those with language problems, refusals, and persistent absence.

Table 32. Final Outcomes for Cases Sent to CATI

Final Status of Case	Fixed	Adaptive
Complete	79.7%	81.1%
CATI complete*	2.5%	4.4%
CAPI complete-occupied	58.0%	50.7%
CAPI complete-vacant/temp occupied	12.7%	19.4%
TQA complete	2.0%	0.8%
Not a housing unit†	4.4%	5.8%
Incomplete/Non-interview††	19.8%	18.9%

Note: For the fixed group, these cases were sent to CAPI with phone numbers and completed by telephone attempt prior to any personal visit attempts. For the adaptive case management group, these are cases completed by CATI interviewers at the centralized CATI facility.
†: Includes Type Cs, cases that are not housing units such as demolished or condemned housing units and businesses for which a status is assigned but no interview is performed.
††: Includes incompletes that were stopped due to maximum contact attempts as well as non-interviews, "Type As" whose households could not be interviewed such as those with language problems, refusals, and persistent absence.

5.4.1 Centralized CATI

There were 609 cases sent to the centralized CATI facility prior to being sent to interviewers for in-person interviewing. These cases had up to three phone numbers, and each case could be called up to 10 times during the two-week field period.

The 609 cases sent to CATI had 1,417 phone numbers. See Table 33 for the final status of phone numbers sent to CATI. Of these numbers, about 11% reached a potential respondent with 1.9% of the phone numbers resulting in a completed interview and 9.4% reaching a person but without completing an interview at that number. The majority (55.8%) of phone numbers were ruled ineligible based on non-contact. For these numbers, the maximum number of unproductive calls (with no prior contact) limit, as set in the call parameters, was reached. We do not know that the number was invalid for that household, but it was ruled ineligible after multiple unproductive call attempts to allow other numbers to be tried. 8.3% of the numbers had a status of unreached when the CATI period ended. This indicates that the numbers were tried, but no contact was made with a person, and the maximum number of unproductive calls for that number was reached. Another 17.8% were recorded as bad numbers - not in service or reached the wrong address or a business. These numbers were confirmed as not reaching the sampled address. About seven percent of numbers were untried. A number could have been untried if the maximum call attempts were exceeded for that case, the maximum refusals were reached (2 per case), or the case was completed with another phone number.

Table 33. Final Status of Phone Numbers Sent to Centralized CATI

Final Status of Phone Number	
Respondent reached (completed)	1.9%
Reached person	9.4%
Unreached	8.3%
Ruled ineligible	55.8%
Message left	0.1%
Bad number	17.8%
Untried	6.9%

There were 6,504 calls made for the 609 cases sent to centralized CATI. See Table 34 for the outcomes of all CATI telephone contact attempts. Of the total calls made in CATI, 0.8% resulted in a completed interview. 8.4% of contact attempts reached a person without completing an interview, including 4.5% of contacts attempts that were refusals or immediate hangups. More than two-fifths (43.7%) of telephone contact attempts reached an answering machine or answering service. One-fifth (20.5%) reached numbers that were disconnected, not in service, or could not be completed as dialed.

Table 34. Outcomes for CATI Contact Attempts

Outcome	Percent of all Telephone First Contact Attempts
Completed interview	0.8%
Contact but no interview	8.4%
Answering machine	43.7%
No answer	14.8%
Busy	4.3%
Disconnected	20.5%
Wrong number	0.4%
Fax	3.5%
Other no telephone contact	3.7%

5.4.2 CAPI Phone Calls

CAPI interviewers were instructed to record all contact attempts in the Contact History Instrument (CHI). They were also instructed to make telephone contact attempts prior to making any personal visit contact attempts for any cases that had telephone numbers. Each phone number was supposed to be attempted twice prior to any personal visit attempt being made.

Of the 552 cases initially sent to CAPI with telephone numbers, 65.4% had an initial contact attempt made by telephone recorded in the CHI. For the remainder of the cases with telephone numbers, the CAPI interviewer recorded a personal visit attempt as the first contact attempt in the CHI.[16]

Table 35. Initial Contact Attempt Type for Cases Sent to CAPI with Phone Numbers

Initial Contact Attempt Type	
Telephone	65.4%
Personal Visit	34.6%

There were 719 telephone contact attempts recorded in the CHI prior to any personal visit attempts for those cases. 1.7% of these contact attempts resulted in a completed interview, and another 6.1% resulted in contact with a person but no completed interview. Over a third (34.5%) reached an answering machine or answering service. Almost a quarter (24.2%) was not answered. 18.6% were disconnected or non-working numbers, and 18.1% were verified as wrong numbers.

Table 36. Outcomes for CAPI Telephone Contact Attempts Made Prior to Personal Visit Contact Attempts

Outcome*	Percent of all Telephone First Contact Attempts†
Completed interview	1.7%
Contact but no interview	6.1%
Answering machine	34.5%
No answer	24.2%
Busy	0.8%
Disconnected	18.6%
Wrong number	18.1%
Fax	1.7%
Other no telephone contact	5.8%

*: Outcomes are not directly comparable to CATI call outcomes due to differences in the systems used to collect contact history data.
†: Numbers do not add to 100% because multiple outcomes could be provided by the interviewer.

[16] During production and in the interviewer debriefings, interviewers reported they did not always make telephone contact attempts prior to personal visit attempts. They explained that this was sometimes because they walked by sample units while performing other interviews, so they took advantage of being near the address, even though they had not made a telephone attempt. In addition, interviewers may not have recorded attempts in the CHI, and, in some cases, may have recorded multiple call attempts in one CHI record.

5.5 Field Training, Supervision, and Monitoring

The 2013 Census Test employed 18 in-person interviewers. These individuals had been recently hired by the Philadelphia Regional Office (RO) to work on the American Housing Survey (AHS). They were selected because they ostensibly had worked on only one Census survey. It was discovered in training, however, that the majority of the interviewers had also worked in previous decennial Census operations.

The office chose in-person interviewers based on an analysis of their CHI compliance (i.e., recording of contact attempts on the laptop) during their work on the AHS and recommendations by the AHS Survey Statisticians Office. Census headquarters staff randomly split interviewers into two groups, fixed (Treatments 1 and 2) and adaptive (Treatments 3 and 4). Interviewers in each group were not meant to be aware of the other group and were separately trained, supervised, and monitored. The RO chose two experienced staff members who served as crew leaders to supervise each group.

5.5.1 Training

The fixed group contained ten in-person interviewers trained over a two-day period from October 31 to November 1. The adaptive group consisted of eight in-person interviewers trained during the same time but at a separate location. Supervisors trained both groups how to use the CAPI instrument, use their laptop's case management system, and fill out CHI. They also performed practice interviews and were trained on key concepts unique to the 2013 Census Test.

Supervisors explained to both groups that the 2013 Census Test would assess new contact strategies with specific cost saving measures important for them to follow. Supervisors reminded in-person interviewers that the 2013 Census Test was about following instructions provided to them through their case management systems. Supervisors stressed the importance of recording their attempts in CHI and informed interviewers they would monitor attempts through a daily report.

After monitoring in-person interviewers' performances in the field, RO and headquarters staff performed a half-day refresher training for all CAPI interviewers. This retraining occurred on November 18, 2013, and each group was instructed separately by their supervisor. The retraining had several goals including (1) gathering questions, concerns, and feedback from interviewers; (2) reviewing key test concepts; (3) making sure procedures were being followed correctly; and (4) reiterating that the objective of the test was to follow instructions rather than complete CAPI interviews at all costs.

The following sections discuss key procedures unique to the 2013 Census Test. These sections discuss four concepts that CAPI interviewers were trained and monitored to do throughout the test. All interviewers were taught and monitored to transmit twice daily on days when they worked and to perform the "Proxy if No Interview" procedure. In addition, the fixed group was

taught and monitored to perform the telephone first procedure, and the adaptive group was instructed and monitored on high priority cases.

5.5.1.1 Transmit Twice Daily Procedure

CAPI interviewers were trained and monitored to transmit once before working, no earlier than 8:00 a.m., in order to receive their daily instructions. Interviewers were also trained to transmit once in the evening after they were done working, no later than 10:00 p.m., in order for the program to run on the most updated contact history and outcome information.

5.5.1.2 "Proxy if No Interview" Procedure

Both fixed and adaptive interviewers were instructed what to do if they observed a "Proxy if No Interview" case on their caselist. "Proxy if No Interview" cases were allowed one final personal visit attempt and were then eligible for proxy attempts to close the case. Training instructed interviewers that "Proxy if No Interview" cases did not need to be prioritized. However, if an interviewer could not obtain an interview with the sample housing unit the day they attempted their final personal visit, they must perform proxy attempts until they collected the necessary information on the housing address *by the end of that day*. The "Proxy if No Interview" procedure was mentioned in three places in the training guide[17] and at one place in each workbook.[18]

5.5.1.3 Fixed Group – Telephone First Procedures

The supervisor trained fixed interviewers to perform two mandatory telephone attempts before in-person visits on all cases with phone numbers. Each address had up to three telephone numbers listed in case management, and interviewers were trained to call all numbers associated with that case two times on two different days at two different times before attempting a personal visit. These "telephone first" procedures were discussed in the interviewer training[19] and located in the training manual[20] and job aid.[21]

5.5.1.4 Adaptive Case Management Group – High Priority Procedure

Adaptive interviewers were trained to attempt all cases marked high priority in case management each day they worked. They were only to attempt a regular case (i.e., not high priority) if it was

[17] Case Management, p. D-13; Walk-Through Proxy Interviews, p. I-3; More Paired Practice Interviews Proxy, p. J-2.
[18] Fixed Workbook, p. 62; Adaptive Workbook, p. 64
[19] p. A-13
[20] p. I-5, I-11-12
[21] Fixed Job Aid, item 4.a.

nearby or had an appointment. This procedure was emphasized in the training[22], the training manual[23], and outlined in the job aid[24].

5.5.2 Supervising Interviewer Performance

For the duration of the 2013 Census Test field period, a daily meeting was conducted with Philadelphia RO supervisors and Census Headquarters staff. The purpose of the Daily Supervisory Meeting was to review in-person interviewer performance, field progress, and discuss supervisor and interviewer issues and experiences. The meeting provided an open forum for the RO to ask questions and headquarters (HQ) to get feedback on the test.

To assist the meeting, the 2013 Census Test Daily Report was created to monitor all in-person interviewers on specific tasks. The report had separate versions for the adaptive case management and fixed groups. The 2013 Census Test Daily Report used ROSCO data to report on transmission compliance, the status of "Proxy if No Interview" cases, the number of cases not checked in (i.e., received), and the number of completed CAPI cases with no contact history information for each interviewer. The adaptive version also contained counts of attempted high priority cases for each interviewer. To supplement this adaptive interviewer report, an HQ staff member created a separate SQL program that identified high priority cases. The fixed version of the 2013 Census Test Daily Report recorded the mode of first contact attempt (telephone or personal visit) of each case to determine if fixed interviewers were following telephone first procedures. (See Appendix A. 2013 Census Test Daily Report (Adaptive Group) for an example adaptive group 2013 Census Test Daily Report.) While discussing this report, discussions focused on the subject of compliance, such as whether or not interviewers transmitted twice daily.

Census HQ staff also prepared a checklist of items to discuss at the meeting. These included discussing and reviewing the 2013 Census Test Daily Report and ROSCO reports such as Cost and Response Management Network (CARMN) reports and the American Community Survey CAPI (ACP) Daily Receipt Report. Time was also set aside to discuss potential issues with systems, discuss inconsistencies between reports, discuss reassignments, and get feedback from supervisors and the RO. See Appendix B. 2013 Census Test Checklist for Daily Supervisory Review, for the 2013 Census Test Checklist for Daily Supervisory Review.

5.5.3 Monitoring Cost and Progress

In traditional Census Bureau surveys, interviewers are trained to complete all CAPI cases without designating the mode of contact, number of contact attempts, or applying special importance to cases. Current surveys and the decennial Census measure in-person interviewer progress based on the number of CAPI interviews completed. For example, cost and progress

[22] Priority Cases in Case Management, p. D-11
[23] Morning Activities, p. C-3
[24] Adaptive Job Aid, item 4a.

reports identify interviewers who do not complete cases at an expected rate. The 2013 Census Test rewarded interviewers who followed instructions over completing cases, a drastically different method of evaluating interviewers. This emphasis on "following the rules" was foreign to both supervisors and interviewers in the 2013 Census Test.

The 2013 Census Test used existing systems, some modified, to collect cost and progress data. ROSCO reports included the ACP Daily Receipt Report, CARMN reports, Final Outcome Code by RO Totals report, and Contact History Instrument report. The ACP Daily Receipt Report listed the workload status of each interviewer by total CAPI workload; cases in supervisory review; number of Type A, B, and C cases; and cases received and not received (i.e., checked-in). The CARMN reports were used by decennial management to monitor survey progress. The other ROSCO reports were used to track interviewer and operation progress.

5.6 Cost

To determine cost savings, we compared productivity across treatments. Because the 2013 Census Test was a site test, the parameters for cost calculations, productivity and total hours, were of more interest than actual cost. While mileage is an important cost factor, it is of less interest in an urban site test.

When conducting the test, in-person interviewers were allowed to work both treatments in a given group (e.g., a fixed interviewer could work both Treatments 1 and 2 in the same day). However, the interviewer then charged his or her time and expenses to the same project code. To separate cost by treatment, we fit the following model, which summarizes each day an interviewer worked:

$$hours = intercept + \#\ successful\ personal\ visits + \#unsuccessful\ personal\ visits + \#telephone\ attempts + miles\ driven$$

The coefficients for each variable in the model were used to parse out the total time for a given interviewer day into its components, or tasks performed by an interviewer. The model uses different variables for three types of visits – successful personal visits, unsuccessful personal visits, and telephone attempts, as each type of visit takes, on average, different amounts of time. For example, it takes longer to visit and complete a successful in-person interview than it does to attempt but not complete an in-person interview. Phone attempts are grouped together because the telephone success rate is lower. Miles driven are used as a proxy for travel time.

After the model is fit, we take the total number of each type of attempt for a given interviewer in a day and the number of miles driven and use the coefficients of the model to calculate the time for each type of attempt. The intercept represents "other" time – or time spent not driving or contacting people (e.g., meeting with a supervisor, planning the workday). We apportioned the intercept equally to each contact attempt. Miles driven represent travel time and were equally apportioned to each personal visit attempt. At that point, we have an estimate for time spent on each contact attempt. To incorporate CATI results, we used a fixed cost rate for each attempt.

Note that the case and attempt numbers may not match other tables in the report because we required payroll data to match contact history data in order to parse out costs. Table 37 shows productivity estimates for each treatment.

Table 37. Productivity Measures for CAPI and CATI by Treatment

			CAPI			Total (CAPI and CATI)	
Treat.	**Cases**	**Attempts**	**Productivity (Cases/Hr.)**	**Hours**	**Attempts/ Hour**	**Productivity (Cases/Hr.)**	**Attempts/Hour**
1	310	1242	0.73	422.27	2.94	0.73	2.94
2	501	1841	0.92	546.49	3.37	0.92	3.37
3	311	841	1.07	291.51	2.88	0.93	6.44
4	496	1143	1.33	371.83	3.07	1.14	7.13

Applying a Wilcoxon sign test, we see all CAPI productivity comparisons are significant at the 0.1 level except Treatments 1 and 2. However, the trend for Treatments 1 and 2 is in the direction we expect.

5.7 Data Quality

In order to assess the data quality for this test, overall completion rates and item non-response rates for the census data items (tenure, relationship, sex, age, Hispanic origin, and race) were calculated.

5.7.1 Completion Rates

Table 38 displays completion rates by treatment for cases not removed prior to data collection with administrative records. Completed cases were considered any with data available from fieldwork. This included completed interviews at occupied housing units, vacant housing units, and cases that were not housing units. Not included in the denominator are cases removed from workload before going to the field.

Completion Rate = (Completes + Vacant + Not Housing Units)/(Total Workload-Cases Removed with AR)

The treatments with administrative records removed before going to the field completion rates of 80.4% for the fixed treatment and 83.1% for the adaptive case management treatment. Treatments with no administrative records removed before fieldwork had completion rates of 76.1% for fixed and 79.4% for adaptive.

Table 38. Completion Rates by Treatment for Cases not Removed Prior to Data Collection

Group	Fixed (n=821)	Adaptive (n=848)
AR removal	80.4%	83.1%
No AR removal	76.1%	79.4%
Total	**77.8%**	**80.8%**

Table 39 displays completion rates by treatment for all cases. Case with data available from fieldwork or from administrative records sources were considered complete. This included completed interviews at occupied housing units, vacant housing units, and cases that were not housing units as well as cases where administrative records (AR) data were available.

Completion Rate = (Completes + Vacant + Not Housing Units + AR Data Available)/(Total Workload)

The treatments with administrative records removed before going to the field had a completion rate of 88.1% for the fixed group and 89.8% for the adaptive case management group for all cases when cases with administrative records data available were included. Treatments with no administrative records removal had completion rates of 84.3% for fixed and 90.3% for adaptive when cases with administrative records data available were included as complete.

Table 39. Completion Rate by Treatment including Cases with Administrative Record Information

Group	Fixed (n=821)	Adaptive (n=746)
AR removal	88.1%	89.8%
No AR removal	84.3%	90.3%
Total	**86.2%**	**90.1%**

Overall, 51% of completed interviews were conducted with a household member, and 12.9% of cases were completed with a proxy. The remaining cases were completed with administrative records data in Treatments 1 and 3. For Treatments 1 and 3 combined, 39% of cases were completed with a household member, 11% with a proxy and 39% using administrative records. For Treatments 2 and 4 combined, 63% of cases were completed with a household member and 15% were completed with a proxy.

Table 40. Completion Rate by Treatment by Type of Complete

Treatment	Cases	Percent Complete,† Household	Percent Complete,† Proxy	Percent Complete, AR Data	Percent Total Completed Cases
1	511	34.1%	15.1%	39.1%	88.1%
2	510	60.6%	15.7%	N/A	76.1%
3	528	43.5%	7.2%	39.4%	89.8%
4	528	65.5%	13.8%	N/A	79.4%
Total	**2077**	**51.0%**	**12.9%**		**83.3%**

†: Complete includes occupied households, vacant housing units, and addresses that are not housing units.
Note: Proxy or household respondent designation was obtained from trace files. Due to reliance on trace files, designation may have been incorrectly recorded.
Note: Vacant housing units and cases that are not housing units are usually completed by proxies but may be closed by the supervisor without proxy verification. A higher percentage of vacant housing units and cases that are not housing units in the adaptive sample area may contribute to a larger percentage of proxy completions.

Table 41. Unit Status by Treatment

Treatment	Cases	Percent Occupied	Percent Vacant	Percent Delete	Complete with AR Data	Percent Total Completed Cases
1	511	38.0%	7.24%	3.5%	39.1%	88.1
2	510	55.1%	14.5%	6.3%	N/A	76.1
3	528	30.1%	15.0%	5.1%	39.4%	89.8
4	528	46.6%	22.4%	10.23%	N/A	79.4
Total	**2077**	**42.4%**	**14.8%**	**6.3%**		

5.7.2 Item Nonresponse

Item nonresponse for Census data items are shown in Table 42 through Table 47.

Table 42. Item Nonresponse Rates for Completed Interviews: Tenure

	Fixed (n=340)	Adaptive (n=298)
AR removal	0.8%	6.3%
No AR removal	2.9%	5.8%
Total	**2.1%**	**6.0%**

Note: Excludes interviews completed with non-knowledgeable proxies who only know housing unit population count.

Table 43. Person Level Item Nonresponse Rates: Relationship

	Fixed (n=539)	Adaptive (n=529)
AR removal	1.0%	1.4%
No AR removal	0.0%	1.6%
Total	**0.6%**	**1.5%**

Note: Relationship is not collected for the householder. This number includes only persons other than householders.

Table 44. Person Level Item Nonresponse Rate: Sex

	Fixed (n=878)	Adaptive (n=853)
AR removal	0.0%	0.9%
No AR removal	0.4%	0.6%
Total	**0.2%**	**0.7%**

Table 45. Person Level Item Nonresponse Rate: Age

	Fixed (n=878)	Adaptive (n=853)
AR removal	0.6%	0.0%
No AR removal	0.9%	0.6%
Total	**0.8%**	**<0.4%**

Table 46. Person Level Item Nonresponse Rate: Hispanic Origin

	Fixed (n=878)	Adaptive (n=853)
AR removal	1.2%	1.1%
No AR removal	1.3%	2.4%
Total	**1.3%**	**1.9%**

Table 47. Person Level Item Nonresponse Rate: Race

	Fixed (n=878)	Adaptive (n=853)
AR removal	1.2%	1.1%
No AR removal	1.1%	3.2%
Total	**1.1%**	**2.3%**

6. Related Evaluations, Experiments, and/or Assessments

The 2013 Census Test Assessment used information from the *Results of the 2013 National Census Contact Test* (Miller and Stewart, forthcoming). The assessment is a pilot test of methods, variants of which will be implemented in the 2014 Census Test and subsequent tests and research projects.

7. Conclusions and Recommendations

7.1 Conclusions

7.1.1 Using Administrative Record Information

The 2013 Census Test provided an initial opportunity to use administrative records to inform removing cases from the NRFU workload. To do this, three administrative records sources and UAA reason codes assigned by the United States Postal Service were used to classify housing units as vacant or occupied. To assign a housing unit as vacant, it was necessary to receive an UAA vacant reason code from the 2013 Census Test mailing. We assigned a vacant status correctly for 67.9% (19/28) and 80.3% (49/61) of the cases in Treatments 2 and 4. For enumeration of occupied housing units, at least one of three administrative records sources needed to have the same validated persons within a source across two years, and we found 42.2% (73/173) of the housing units have a household size based on administrative records that is smaller than the NRFU interview. These rules might have led to misclassifying some cases.

7.1.2 Reducing Contact Attempts

The 2013 Census Test examined a reduction in the number of maximum personal visit attempts. In the 2010 decennial Census, enumerators were permitted up to six contact attempts for each case, while the 2013 Census Test had a maximum of three personal visit attempts in Treatments 1, 2, and 3. In Treatment 4, the presence or absence of administrative records suitable to enumerate the household determined the level of contact effort, either one or three maximum personal visit attempts. In total 345 of 1,642 cases were incomplete (i.e., stopped in the field due to maximum attempts). Of these, 100 could be enumerated with administrative records while 245 had no data. There is little information on these 245 cases, but in-person interviewers recorded experiencing a refusal or reluctance with 29.0% of the cases, while 68.6% had only non-contacts.[25] Although we cannot generalize to a broad population, in this test we saw a large percentage of cases with no data that could be candidates for imputation.

7.1.3 Implementing an Adaptive Design Case Management Strategy

The 2013 Census Test demanded that CAPI interviewers follow strict contact procedures. These included transmitting data to and from their laptops twice each day that they worked, filling out the contact history instrument, and performing proxy interviews after their final personal visit

[25] The remainder of cases experienced language barriers or other issues associated with contacting the household but not refusals or reluctance.

contact attempt. Interviewers in the adaptive case management group were also instructed to attempt high priority cases first.

In general, CAPI interviewers followed transmission rules. Over 85 percent of transmissions were compliant on days interviewers worked, occurring between 6:00 a.m. and midnight. Interviewers' transmission compliance did not increase over the CAPI field period. Interviewers were less compliant in performing proxy interviews after their final personal visit and working all high priority cases. Of cases on their final personal visit attempt, 28.46% were attempted only one time and left uncompleted. Some interviewers had a majority of their incomplete cases stopped after only one personal visit attempt, indicating they possibly did not understand the "Proxy if no Interview" procedure. Interviewers did not attempt all high priority cases even though they worked regular cases nearly 50 percent of days with compliant transmissions (defined between 6:00 a.m. and midnight).

To test propensity score predictions made by the Case Prioritization Model, we compared contact and completion rates between regular and high priority cases in Treatments 3 and 4. While contact rates are similar between regular and high priority cases in both treatments, completion rates are lower for high priority cases compared to regular cases in both Treatments 3 and 4.

These findings suggest that some of the procedures in the test were implemented well, while others did not achieve good compliance. Implementation failures make it more difficult to judge propensity score predictions and experimental manipulations.

7.1.4 Employing Telephone Calls

When phone numbers are available that belong to a household member of a sample address, phone interviewing is clearly a less expensive alternative to personal visits by an in-person interviewer. However, commercial vendors performed the matching of phone numbers to addresses for this study, and the quality of matching was problematic.

Only 2.5% of fixed and 4.4% of adaptive cases that had telephone numbers and were not removed with administrative records were completed by telephone in the centralized calling center or, for fixed cases, by telephone prior to any personal visit attempts being made. Two-fifths of telephone contact attempts by in-person interviewers in the fixed group resulted in no telephone contact, such as reaching a number ringing with no answer or being disconnected. Over one-third of calls by in-person inteviewers reached an answering machine. Over one-third of call attempts from the CATI facility did not result in telephone contact. However, more than two-fifths of CATI call attempts reached an answering machine. The majority of the telephone numbers sent to the centralized CATI facility did not reach anyone after a maximum number of calls. Over half of the telephone numbers sent to CATI were ruled ineligible based on non-contact. The presence of many answering machine cases suggests that calling does provide potentially useful information about occupancy. The number of cases ruled "ineligible" suggests that there is potential for obtaining contact with different calling rules.

The CATI and CAPI call outcomes were not directly comparable for this test due to differences in the systems for recording contact history information. These differences made it difficult to compare results of the two calling regimes.

7.1.5 Field Training, Supervision and Monitoring

The 2013 Census Test fixed group had ten interviewers trained over a two-day period from October 31 to November 1. The adaptive group had eight interviewers trained during the same time at a separate location. Supervisors trained both groups how to use the CAPI instrument, use their laptop's case management system, and fill out CHI. They also were trained on key concepts. After monitoring interviewers' performances in the field, RO and headquarters staff performed a half-day refresher training for all interviewers.

Census Bureau trainings traditionally teach in-person interviewers to perform specific types of interviews (e.g., occupied, vacant) and account for specific types of situations (e.g., language barrier), rather than follow specific instructions such as performing proxy attempts after a final personal visit attempt. Both supervisors and interviewers said that more exercises relating to such contact strategies would have helped them understand unique procedures.

In traditional decennial NRFU, enumerators are trained to complete all cases without designating the mode of contact, number of contact attempts, or applying special importance to cases. The 2013 Census Test emphasized following instructions over completing cases, a very different method of evaluating in-person interviewers. Furthermore, the 2013 Census Test did not set "progress goals" because it was difficult to estimate how long the field period would take with cases being stopped in the field after varied numbers of contact attempts, and because the goals might interfere with implementation of the contact rules under study.

The different conditions in this study made it difficult to train and supervise CAPI interviewer performance.

7.1.6 Cost

In general, the treatments where administrative records were removed before fieldwork had lower productivity than other treatments (Treatments 1 v. 2 and 3 v. 4) because easier cases were removed, and the cases left were more dispersed. However, interviewers charged fewer hours in treatments where administrative records were removed before fieldwork, leading to cost savings albeit lower productivity. An adaptive approach to case management also resulted in a jump in productivity (Treatments 1 versus 3 and 2 versus 4). The results suggest that, even with the difficulties previously described, the use of administrative records to removed workload and the implementation of an adaptive design case management approach have potential to reduce costs.

7.1.7 Data Quality

Completion rates for cases not removed before fieldwork were comparable between the fixed (77.8%) and adaptive (80.8%) groups as were completion rates that counted all cases whose status could be determined with administrative records as a complete (86.2% for fixed and 90.1% for adaptive). Overall, 13% of cases were completed with a proxy interview. Item nonresponse rates were low across all data items.

7.2 Recommendations

The following recommendations come out of the results and conclusions presented in this assessment.

7.2.1 Administrative Records

Research different sources and criteria for inferring from administrative records the occupancy status and population size of housing units.

Consider using additional sources to designate vacant housing units. Examination of sample cases found to be vacant in the field but not identified as such from the USPS UAA "vacant" code suggests that it may be wise to consider using additional UAA codes (e.g., "attempted - not known") when designating housing units as vacant.

Consider relaxing some rules used to match persons with occupied housing units. Examination of sample cases found to be occupied in the field but not identified as such through administrative records suggests that it is useful to consider relaxing the criteria requiring a two-year match of all persons within an administrative records source in order to designate a unit as occupied. Designating cases as occupied with at least one PIK in agreement across multiple years (instead of all PIKs) may be sufficient. Alternatively, we could only rely on one year of data in order to use the administrative records source as a justification for removing occupied cases from the NRFU workload. Since the rule requiring persons to be found at an address for two consecutive years in administrative records led to undercounts when compared with information obtained in field interviews, we should examine allowing persons associated with the unit in records for one year to be part of the household count.

7.2.2 Reduced Contact Attempts

Research how to obtain proxy interviews and judge their quality.

Research the timing of proxy attempts. In the decennial Census, proxies played a crucial role. They verified vacant housing units and addresses that should be deleted as well as enumerated housing units after the final contact attempt. Yet, interview attempts with proxies often do not performed strategically. In-person interviewers in the 2013 Census Test felt a "knowledgeable" proxy should be interviewed when the opportunity presented itself rather than waiting until a case became a high priority or was on its last attempt. Interviewers mentioned that they

sometimes got interviews with a knowledgeable proxy and later needed a proxy interview for another house on the same block, but the knowledgeable proxy was unavailable. It may be useful to gather information from a proxy respondent prior to exhausting all contact attempts with a housing unit and use that information if no interview is obtained at the address.

Future trainings should stipulate how "knowledgeable" a proxy respondent needs to be on occupied housing units. During debriefings, several CAPI interviewers were unsure whether they needed to obtain all the questionnaire information (e.g., birthdate, sex, and race) on a roster or just a population count. These interviewers suggested that they made multiple proxy attempts in an effort to find a proxy respondent who could provide them all information rather than obtaining a population count from the first available proxy.

Assess the use of administrative records and methods of imputation for cases with no data after interviewing is complete.

Evaluate the tradeoffs between data quality gathered from interviews, administrative records, and imputation. No data were obtained for a noteworthy number of cases in this test due to a reduced maximum number of contact attempts. Thus, it seems advantageous to research methods for increasing data production (proxy interviews, administrative records, and imputation) and evaluate their quality.

7.2.3 Adaptive Case Management

Design case management and interviewing systems that are "user friendly" for enumerators.

Instructions for handing cases should be usability tested. The 2013 Census Test case management displays for CAPI interviewers – indicating when a case was high priority and when it was to receive its final contact – were modifications to the case management screens employed in the American Community Survey. The modifications involved adding exclamation points and icons to case management screens. These symbols were not usability tested prior to the study due to lack of time and resources. Features like this should be tested to ensure they are prominent and there is no confusion regarding their meaning.

Future trainings should consider spending time on instructing enumerators how to plan their day. The 2013 Census Test briefly instructed in-person interviewers on how to plan their day both during training[26] and in their manual.[27] However, debriefings uncovered that interviewers did not approach day planning in the same way. Some interviewers reported that they bounced around specific neighborhoods, attempting to complete high priority cases and appointments. Another interviewer described how she rigidly planned each day's cases, recording which were

[26] Adaptive Design Training Guide, Case Management, p. D-22

[27] Adaptive Design Field Representative Manual, Planning Route, p. 3-25 to 3-26 and Appendix C, CAPI, Prepare to Interview, p. C-4

high priority and "Proxy if No Interview" cases. Prescriptions for day planning may be helpful to interviewers.

Research alternative methods for determining the response propensity of sample units.

Determine how to develop Day 0 models for future tests. The 2013 Census Test Day 0 Model relied on information from 2010 NRFU retrieved from the LCAT database. Future tests may not be able to rely on these data, as they are not available for cases that self-responded to the 2010 Census. For tests with mailout components, models should take into account households that did not participate in past non-response follow ups, such as cases that self-responded in past decennials and new housing units. Furthermore, the LCAT database may become less relevant over time as households move and neighborhoods change. Future research could also investigate 2-way interactions for variables as well as non-parametric models.

Determine how to treat cases that are in "supervisory review." During the 2013 Census Test, a case went into supervisory review status if it was a Type C (i.e., not a housing unit) or it was a duplicate case. Duplicate cases have already been received or checked in, but another version of the case is transmitted to ROSCO. The supervisor must then decide what to do with the new version of the case - overwrite the status of the previous case or delete the new case. The test did not look at supervisory review data. In certain instances, this meant the program's instructions may have been ignored. For example, the program may have instructed that a case be removed from a laptop, but a supervisor may have decided to accept that same case as a Type C in supervisory review. Depending on the timing, the case could show up on a report as an incomplete case and then as a Type C. The business rules for handling cases in supervisory review and the business rules for when versions of a case can or cannot override a previous version of a case need to be carefully considered in future designs.

Determine which contact history records to include in models. The 2013 Census Test did not include proxy contact attempts in the Case Prioritization Model, as contact attempts with households exhibit different patterns than contact attempts with proxy respondents. For example, whether or not a sample unit has been previously contacted may exhibit a positive relationship with completing a household interview but no relationship at all with completing a proxy interview. It is likely that these two forms of contact attempts are best modeled using different parameters. Due to CHI pathing, certain contact history variables in the 2013 Census Test were only available when in-person interviewers said they were performing interviews with sample unit members rather than proxy members. Future designs must consider what to do with proxy contact attempts, especially during decennial Census tests, as proxy interviews make up a large portion of occupied housing units. Future designs must also consider how to deal with telephone contact attempts, as they exhibit different patterns than personal contact attempts.

Determine which outcome codes to include as a completed CAPI interview in models. The Case Prioritization Model for the 2013 Census Test considered six types of CAPI cases to be "completed": occupied units, sufficient partials, cases completed using TQA, vacant units, and

temporarily occupied units. However, vacant and demolished units exhibit different contact patterns than occupied housing units due to different procedures. For example, enumerators in NRFU are trained to look for a proxy respondent when a case is likely not a housing unit, and contact history instruments differ across suspected vacant, deleted, and occupied units. It is likely that significant predictors of vacant or demolished housing units differ than significant predictors of habited housing units. Further research should determine how best to handle vacant and demolished housing units under an adaptive design. Mixed mode data collections bring similar challenges, and decisions must be made up front with what to do with late returns completed via a telephone mode, internet, or paper.

Consider how to assign priority scores to cases with no contacts. The 2013 Census Test assigned a propensity score to all cases with contact history information, while cases without at least one recorded contact attempt were assigned their initial propensity scores. Research should compare the distributions of Day 0 and future propensity scores.

Develop protocols to test programs used in to execute adaptive case management procedures.

If test data are not available with which to construct a program, future tests should use a preliminary systems test to generate data. The first program generated for the 2013 Census Test broke several times during the first systems test. Rather than attempt to run the code, it may have been advantageous to populate the instructions passed to the operational control system by hand and use the data created during the systems test to build the program.

Set up processes to ensure that systems test scenarios are keyed as written. During the first systems test, several scenarios were not keyed correctly or at the correct time, resulting in unanticipated instructions sent to laptops. In these instances, researchers had to access ROSCO and study lines of business rule code to uncover the error. Many mis-keyed scenarios had to be further modified to test the systems test scenarios correctly.

Examine a diverse set of scenarios for systems tests. The 2013 Census Test keyed nearly 150 CAPI scenarios over the course of two systems tests. These scenarios included many complicated systems interactions like case reassignments and CATI recycles. However, on the first day of the field period, the program broke because the study team did not test what would happen if an in-person interviewer did not do any work on the first day of the field period. While not all scenarios may be anticipated, systems tests should include a diversity of keyers and scenarios in an attempt to mitigate any problems during production.

Research how to make daily case assignments for enumerators, including the size of the daily workload and its geographic composition.

Future tests and analysis should measure whether case prioritization causes enumerators to attempt more cases per day. The Philadelphia RO suggested that interviewers, on average, could

attempt about six cases each day, so adaptive interviewers were assigned seven high priority cases each day. Interviewers reported they felt pressured to attempt these cases each day, and one remarked in the debriefing questionnaire that regular cases were not as stressful. Future tests and analysis should disentangle if adaptive case management procedures lead to efficiency because CAPI interviewers attempt cases that are more likely to be completed each day or because interviewers are taught and supervised to attempt more cases in general.

Consider a geographic component to modeling priority cases. The 2013 Census Test did not take into account the proximity of cases to each other or to the CAPI interviewer when prioritizing cases. The non-geographic consideration of high priority cases assignment was noted in the questionnaire debriefing. Wrote one interviewer, "Sometimes I had to jump over cases to ensure that I visited all [high] priority cases first during my trips." Analysis of the 2010 Census suggests that clustering cases may reduce NRFU mileage and hours (Hatipoglu, 2014).

7.2.4 Telephone Calls

Explore alternative ways to match telephone numbers with sample addresses and rank order them for likelihood of contact.

Consider how research on the ACS telephone frame may improve telephone-address matches. The results of telephone operations in the 2013 Census Test revealed that many numbers matched to sample addresses by the third party data providers were not accurate. Recent research on the ACS telephone frame suggests that alternative approaches for identifying the telephone numbers likely to be correctly linked to the sample address may be useful. These approaches may be applicable in the decennial Census.

Consider how research on the telephone frame for the ACS may improve the productivity of telephone contacts. Both CATI and CAPI telephone calls in the 2013 Census Test were unproductive in garnering completed interviews. Recent research on prioritizing telephone numbers for calling in the ACS may improve the productivity of contact attempts in future decennial Census tests.

Explore alternative methods of determining the quality of telephone numbers matched to addresses.

Make use of match-quality measures and paradata to prioritize telephone numbers. Propensity scores calculated for each phone number-address combination should be considered, along with other paradata related to the address, to decide the value in attempting a call using a particular phone number for a particular address.

Examine alternative CATI calling algorithms to enumerate sample households.

Consider how research on ACS calling algorithms may improve the productivity of telephone contact attempts. Research on ACS CATI outcome codes suggests that altering calling

algorithms can improve productivity. Lessons from this recent research may be applicable to future decennial Census tests.

Consider other changes in calling procedures. For the 2013 Census Test, CATI call parameters attempted numbers a maximum number of times for each case. Future tests should consider adjusting call parameters so all phone numbers are attempted a minimum number of times instead of a maximum number of times for a case.

Examine the tradeoffs and approach to "telephone first" procedures in the field.

Future tests should consider the costs and benefits of the contact procedures used in this test. For several reasons, in-person interviewers did not always follow telephone first procedures. In some instances, the telephone first rule did not make sense to the interviewer. One interviewer stated he discovered a vacant housing unit but was forced to conduct the mandatory telephone calls in his car before he could proceed with obtaining a proxy interview to confirm that the unit was vacant. This situation was also applicable to sample units where interviewers encountered residents of the household while visiting nearby houses. A balance must be struck between the cost savings associated with the prescribed contact procedure and the realities of interviewing situations.

Future tests of the telephone first procedure should consider training in-person interviewers to make all telephone calls to a geographic area first, not just all calls on a case. Preliminary research suggests that a large portion of the costs of NRFU enumerators can be attributed to the time and mileage associated with driving to and from cases (Hatipoglu, 2014). CAPI interviewers remarked in debriefings that they would sometimes attempt personal visits at neighborhoods before they had made telephone calls to each address in the neighborhood. Efficiency may be gained by ensuring interviewers perform telephone calls for entire neighborhoods, rather than just cases, before traveling to the neighborhood.

Standardize the metrics for CAPI and CATI call outcomes.

Develop standard outcome codes for use in both CATI and CAPI telephone contact efforts. The lack of telephone outcome code standardization in the 2013 Census Test made it difficult to compare the results of efforts across CATI and CAPI treatments. Future tests that compare CATI and CAPI telephone operations should build outcome codes that allow the comparison of CATI and CAPI call outcomes.

7.2.5 Training, Supervision, and Monitoring

Achieve, through training, monitoring or automation, higher in-person interviewer compliance with case management and interviewing procedures.

Future tests should automate transmissions. During the 2013 Census Test, CAPI interviewers transmitted before and after working and between 6:00 a.m. and midnight on 85.5% of the days

they worked. The rest were days interviewers forgot to transmit before or after they began work or did not transmit at all. If an interviewer did not transmit at the end of a workday, the propensity model would not be updated with the most recent contact attempts for any worked cases, causing those cases' propensity scores to be based on previous contact attempts. If an interviewer did not transmit before they started working for the day, it meant that the interviewer did not attempt the updated high priority cases that day and may have allowed interviewers to receive additional contact attempts for cases that had reached their personal visit attempt maximum. Future tests must stress the importance of transmitting data correctly and automate such transmissions whenever possible.

CAPI interviewer training should emphasize how to fill out contact attempt records. Although trainings emphasized the importance of completing the CHI, some CAPI interviewers mentioned they were unaware they should have done so for certain attempts. For example, the supervisor from the fixed group reported that some interviewers did not record telephone contact attempts in CHI, and one interviewer stated he recorded telephone calls in case notes instead.

When testing the outcomes of CAPI telephone procedures, systems must accurately record contact attempts. There was evidence that CAPI interviewers did not accurately record their telephone contact attempts during the 2013 Census Test. In order to measure the performance of a telephone call procedure, researchers need accurate and complete data about phone attempts. Future tests should emphasize in trainings when and how telephone contacts are to be recorded. When possible, tests should improve systems to record telephone contact attempts without relying on interviewers.

Future tests should develop trainings that emphasize prescribed contact instructions. In general, Census Bureau trainings teach CAPI interviewers to perform specific types of interviews (e.g., occupied, vacant) and account for specific types of situations (e.g., language barrier), rather than follow specific instructions such as performing multiple proxy attempts after a final personal visit attempt. Both supervisors and interviewers said that more exercises relating to such contact strategies would have helped them to understand procedures unique to the 2013 Census Test.

Future tests should develop trainings and practice cases that emphasize "Proxy if No Interview" instructions. During the Daily Supervisory Meeting with Census Bureau headquarters, supervisors, and other Philadelphia RO staff, it became apparent that staff and interviewers were confused regarding "Proxy if No Interview" procedures. In current surveys, proxies are not allowed or utilized on occupied housing units, and this new concept required more explanation to supervisors. Some interviewers presumed they had to get an interview, from either the household or proxy, on the first day they saw the "Proxy if No Interview" icon on their case management rather than waiting until the case was a high priority or it made sense to visit. Supervisors stressed the need for more training on this concept in debriefings. They suggested introducing

this concept early in training in a self-study and adding more practice "Proxy if No Interview" cases and role-playing to training.[28]

Future tests should develop trainings and practice cases that better prepare enumerators to make "cold" telephone calls. Some in-person interviewers did not understand "telephone first" procedures. Specifically, interviewers were confused as to the number of telephone calls they needed to make. One interviewer reported in the debriefing questionnaire, "Didn't realize that [I] had to make 2 phone calls on two different days at 2 different times." Future tests' classroom training should practice telephone procedures and recording telephone outcomes.

Future tests should spend adequate time editing and testing reports that monitor CAPI interviewer performance and progress. Due to time constraints, the 2013 Census Test Daily report was not tested. As a result, it was necessary to edit the report throughout the test to produce accurate counts. Testing reports *ad hoc* and during systems tests will help to uncover report errors before production, and editing reports prior to data collection allows for more usable reports.

Future tests should spend ample time training supervisors how to read reports and manage with reports. Due to time constraints, supervisors and field staff were not trained to use the 2013 Census Test Daily Report. Instead, time was devoted to discussing how to read and use the report during daily meetings. During their debriefing, supervisors stated they would have benefitted from a training session on how to read, analyze, and view the 2013 Census Test Daily report prior to the beginning of field operations.

Future tests should create reports that allow supervisors to "drill down" to specific problematic cases and days so they can discuss these instances with enumerators. The 2013 Census Test Daily Report conveyed aggregated case information for each CAPI interviewer. For example, it reported on the number of high priority cases not attempted each day per interviewer and the number of cases that were removed from an interviewer the previous day. Throughout the course of the test, it became clear that using these reports to manage necessitated knowing which cases had not been attempted or were removed so that supervisors could talk to interviewers about specific cases. Future tests should consider "drilling down" on cases that are incomplete (i.e., stopped due to maximum attempts) or high priority as well as specific transmissions and attempts per case.

Future tests should consider how to manage reassignments, a traditionally "ad hoc" process, in a structured data collection environment. The 2013 Census Test relied on CAPI interviewers following structured procedures. However, supervisors were not instructed how to reassign cases in supervisory review, which often caused deviations from this structure. Future research should study how reassignments are made, and future tests should consider automating reassignments where possible.

[28] Proxy practice scenarios were located in the Adaptive Workbook, p. 62, Scenarios #5 & #6.

Develop cost and progress reports suited to complex experimental designs.

Future tests should separate costs across treatments. The 2013 Census Test relied on cost and progress data from existing survey systems. As a result, it was difficult to separate costs across treatments. For example, TQA was used across all treatments, and it was reported in the aggregate. In another example, Treatment 2 in the 2013 Census Test allowed cases that could be enumerated with administrative records up to three personal visit attempts, while Treatment 4 allowed only one personal visit attempt, and Treatments 1 and 3 removed these cases entirely. Cost and progress reports used to compare treatments must adequately report such costs so individuals monitoring the study do not make improper assumptions during data collection like reporting "lower costs" on a treatment that has a higher cost per attempt or cost per case by design.

Future tests must consider dependent variables when reporting on progress across treatments. Treatments 2 and 4 allowed cases that could be enumerated using administrative records to go to the field. In these treatments, some incomplete cases could be enumerated with administrative information while others had no information and might be candidates for imputation. However, all incomplete CAPI cases stopped due to maximum attempts showed up as "completes" on monitoring reports, which caused some confusion.

Develop supervisory methods and CAPI interviewer incentives that promote compliance with case management rules, interview completion, and high data quality.

Future decennial tests should develop "progress goals" that take into account incomplete cases stopped due to maximum attempts. The 2013 Census Test did not set progress goals because it was difficult to estimate progress on incomplete cases stopped in the field after maximum attempts, a new concept. There was also concern that such goals would incentivize supervisors and interviewers to move quickly through cases and disregard specific instructions. Future tests should consider how to develop progress goals that account for incomplete cases. One suggestion is to simulate progress goals under specific constraints, such as a maximum number of personal visit attempts on specific types of cases.

Future tests should develop enumerator standards that balance completed CAPI interviews, data quality, and the following of instructions. In current surveys and the decennial tests, ROs establish progress goals that measure CAPI interviewer productivity. In the 2013 Census Test, supervisors monitored interviewers on how well they followed instructions rather than how many CAPI interviews they completed. Philadelphia RO staff stated the need to establish new standards that measure interviewers' abilities to follow instructions. Interviewer debriefings also suggested the need for this balance. Said one interviewer, "Overall this was a very pleasant and positive experience. My only suggestion would be to overemphasize [that] the purpose/goal of the test is to document attempts and not the completion of an interview." For supervisors, this means a more "hands on" approach to managing their staff and requires them to contact interviewers when they do not comply with prescribed procedures. It also necessitates incentives

for interviewers who both follow instructions and complete interviews. Staff mentioned that rewarding such interviewers with more cases could be effective if supervisors can identify people who meet production goals and are compliant with procedures.

8. Knowledge Management Recommendations

Although the team responsible for this test did not receive formal knowledge management recommendations, the assessment touched on the following:

Table 48. Knowledge Management Recommendation Touched on by 2013 Census Test

Number	Document Title	Project Level Disposition Code	Description
2	*Change and the 2020 Census: Not Whether But How*	No Disposition Code Available†	Focus research and development efforts on priority topic areas to achieve a lower cost, high-quality 2020 Census including using multiple modes for respondent convenience and data quality.
3	*Change and the 2020 Census: Not Whether But How*	No Disposition Code Available†	Focus research and development efforts on priority topic areas to achieve lower cost, high-quality 2020 Census including use of administrative records to supplement and improve a variety of operations.
18	*Census 2010: Final Report to Congress*	No Disposition Code Available†	Explore alternative approaches for conducting the 2020 Census that include utilizing administrative records.
818	*2010 Census Match Study Evaluation Report*	No Disposition Code Available†	While the quality and coverage of administrative records suggests that they can be utilized in decennial Census operations, the quality is not high enough and the coverage not expansive enough to replace a traditional decennial Census.
820	*2010 Census Match Study Evaluation Report*	Pursued	To reduce costs, administrative records can assist in determining housing unit and occupancy status.
821	*2010 Census Match Study Evaluation Report*	Pursued	Further research should be conducted on population count differences between administrative records and the 2010 Census.

†: Awaiting feedback from owner or additional review by the ADC Review Board

9. Acknowledgements

This evaluation, analysis, and assessment report is the product of many contributors. Gina Walejko, Andrew Keller, Gianna Dusch and Peter Miller are the main authors of the report. The work could not

have been completed without the significant contributions of these colleagues: Tamara Adams, Fernando Armstrong, Louis Avenilla, Karen Bagwell, Stephanie Coffey, Jaya Damineni, Kevin Deardorff, Dennis Donahue, Chandra Erdman, Bryn Johnson, Susanne Johnson, Ganesan Kakkan, Scott Konicki, Alfred Meier, Caleb Miller, Jaclina Mommsen, Tom Mule, Shadana Myers, Travis Pape, Joseph Quartullo, David Sheppard, Melissa Therrien, James Wagner, Danielle Williams and Derek Young.

10.References

(2012), "Business Plan for the 2020 Census in Support of the FY 2013 Budget Submission," Version 2.0, June 27, 2013.

Coffey, Stephanie, Benjamin Reist and Michael White (2013), "Monitoring Methods for Adaptive Design in the National Survey of College Graduates (NSCG)," *JSM Proceedings*, Montreal, Canada: American Statistical Association.

Groves, Robert and Mick Couper (1998), *Nonresponse in Household Interviewer Surveys*, New York: Wiley.

Groves, Robert and Steven Heeringa (2006), "Responsive design for household surveys: tools for actively controlling survey errors and costs," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 169(Part 3): 439-457.

Hatipoglu, Cem (2014), "Mileage/Cost Reduction Opportunity Assessment in Enumerator Nonresponse Followup Operations," February 2014.

Miller, Peter (2013), "Adaptive Design at the U.S. Census Bureau," Workshop Advances in Adaptive and Responsive Survey Design, Statistics Netherlands, Heerlen, Netherlands, December 10, 2013.

Miller, Caleb and Tim Steward (forthcoming), "Results of the 2013 National Census Contact Test."

Rastogi, Sonya and Amy O'Hara (2012), "2010 Census Match Study," 2010 Census Program for Evaluations and Experiments, No. 247, November 16, 2012.

Sheppard, David (2014). "Phone Contact Frame Data Selection Options," CARRA Working Paper CARRAMEN 05-14, April 14, 2014

Walker, Shelley, Susannah Winder, Geoff Jackson and Sarah Heimel (2012), "2010 Census Nonresponse Followup Operations Assessment," 2010 Census Planning Memoranda Series, No. 190, April 30, 2012.

11. Appendix A. 2013 Census Test Daily Report (Adaptive Group)

Team 80		Prev Day's Transmissions		Telephone Cases (cumulative)			PV/Proxy Cases Checked-In (cumulative)				Cases Not Checked In (daily)					PV/ Proxy Cases Open > 3 Days (daily)	Completed Cases w/ No CHI Records (cum.)
Report Date	**FR Code**	**1st Trans Time**	**Last Trans Time**	**Count**	**1st Attempt T**	**1st Attempt PV**	**Count**	**Stopped**	**Completed**	**LMR**	**Count**	**High Priority**	**Reg Priority**	**PV/ Proxy**	**Stop Work**		
11/16/2013 (previous day)																	
11/17/2013 (today)																	

12. Appendix B. 2013 Census Test Checklist for Daily Supervisory Review

Below is the suggest list of items for review prior to the Daily Supervisory Review

☐ Run ROSCO Reports:
- ☐ ACP Daily Receipt Report by interviewer name
- ☐ Supervisory Review, Type A,B,Cs
- ☐ CARMN Reports

Review tasks of each interviewer if they worked
- ☐ Did they transmit twice?
- ☐ Did they complete a CHI for each attempt on their cases?
- ☐ Did they complete entries for daily hours and expenses on WebFRED?

☐ Review 2013 Census Test Daily Report (upon receipt)

For Adaptive Group (81)
- ☐ Review columns for transmission records, (2 times, when?, did they work?)
- ☐ High priority case attempts (each high priority case attempted, regular case attempted)
- ☐ PV/Proxy cases, (review amount of cases knowing stop work and LMRs may affect amount)
- ☐ Cases not Checked In, (shows workload per interviewer for not-checked in)
- ☐ PV/Proxy over 3 days, (compare with high priority wkld),
- ☐ Completed Cases w/o CHI (hopefully rare, info recorded with interviewer?)

For Fixed Group (80)
- ☐ Review Columns for Transmission records, (2 times, when?, did they work?)
- ☐ Telephone attempts (1st attempt should be telephone)
- ☐ PV/Proxy cases (review amount of cases knowing stop work and LMRs may affect amount)
- ☐ Cases not Checked In (shows workload per interviewer for not-checked in, compare to previous day's count)
- ☐ PV/Proxy over 3 days (chk for difficult of obtaining PV)
- ☐ Completed Cases w/o CHI (hopefully rare, info recorded with interviewer?)

☐ Record and share potential issues with the condition regarding:
- ☐ System issues with ROSCO, WebFRED, Laptop, Case Management
- ☐ Report inconsistencies between the 2013 Census Test Daily Report and ROSCO Reports
- ☐ Feedback from the interviewers regarding e.g. workload, procedures
- ☐ Case Reassignments e.g. Language, procedural, sup review

NOTES:

13. Appendix C. Acronyms and Abbreviations

Acronym/Abbreviation	Definition
#	
A	
ACP	American Community Survey Computer Assisted Personal Interview
ACS	American Community Survey
AHS	American Housing Survey
AR	Administrative Records
B	
C	
CAD	Center for Adaptive Design
CARMN	Cost and Response Management Network
CARRA	Center for Administrative Records Research and Applications
CAPI	Computer Assisted Personal Interview
CATI	Computer Assisted Telephone Interview
CF	Contact Frame
CHI	Contact History Instrument
CLD	Crew Leader District
CMS	Center for Medicare and Medical Services
CUF	Census Unedited File
D	
DPV	Delivery Point Validation
DSF	Delivery Sequence File
E	
F	
FCC	Federal Communications Commission
FR	Field Representative
FY	Fiscal Year
G	
H	
HQ	Headquarters
HU	Housing Unit
I	
IRS	Internal Revenue Service
J	

K

L

LCAT	Life Cycle Analysis Team

M

MAF	Master Address File
MAFID	Master Address File Identification
MCS	Master Control System
MEDB	Medicare Enrollment Database
DTDB	MAF/ TIGER Database

N

NANPA	North American Numbering Plan Administration
NPC	National Processing Center
NRFU	Nonresponse Follow-up
NSCG	National Survey of College Graduates

O

OMB	Office of Management and Budget

P

PIK	Protected Identification Key
PV	Personal Visit

Q

R

RO	Regional Office
ROSCO	Regional Office Survey Control Operations

S

SQL	Structured Query Language

T

TCC	Tucson Contact Center
TeNTAP	Telephone Number Type Assignment Process
TIGER	Topologically Integrated Geographic Encoding and Referencing
TFC	Targus Federal Consumer
TQA	Telephone Questionnaire Assistance

U

UAA	Undeliverable-as-Addressed
USPS	United States Postal Service
UTS	Unified Tracking System

V

W

WebCATI	Web Computer Assisted Telephone Interviewing
WebFRED	Web-based Field Representative Employee Data

X

Y

Z