UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

2020 CENSUS PROGRAM INTERNAL MEMORANDUM SERIES: 2019.12.i

Date: April 12, 2019

MEMORANDUM FOR: The Record

From: Deborah M. Stempowski **(signed April 12, 2019)**
Chief, Decennial Census Management Division

Subject: 2020 Census Evaluation: Administrative Record Dual System Estimation Study Plan

Contact: Jennifer Reichert
Decennial Census Management Division
301-763-4298
jennifer.w.reichert@census.gov

This memorandum releases the final version of the 2020 Census Evaluation: Administrative Record Dual System Estimation Study Plan, which is part of the 2020 Census Program for Evaluations and Experiments (CPEX). For specific content related questions, you may also contact the authors:

Thomas Mule
Decennial Statistical Studies Division
301-763-8322
vincent.t.mule.jr@census.gov

Andrew Keller
Decennial Statistical Studies Division
301-763-9308
andrew.d.keller@census.gov

Scott Konicki
Decennial Statistical Studies Division
301-763-4292
scott.m.konicki@census.gov

2020 Census Evaluation

Administrative Record Dual System Estimation Study Plan

Thomas Mule, Decennial Statistical Studies Division
Andrew Keller, Decennial Statistical Studies Division
Scott Konicki, Decennial Statistical Studies Division

April 12, 2019
Version 1.3

Page intentionally left blank.

Table of Contents

Page intentionally left blank.

I. Introduction

This evaluation expands upon the innovation of utilizing administrative records and third-party data for the 2020 Census Address Canvassing and Nonresponse Followup operations. The Census Bureau also has a long history of using administrative records of birth, deaths, and other information to produce Demographic Analysis (DA) coverage estimates (Robinson et al. 1993, Robinson et al. 2002, and U.S. Census Bureau 2012). This evaluation expands this innovation by attempting to see if administrative records and third-party data could be used to produce capture-recapture coverage estimates. Since 1980, the Census Bureau has produced capture-recapture coverage estimates by conducting an independent post-enumeration survey and using dual system estimation approaches.

Because of budget constraints, the 2020 Census Post-Enumeration Survey (PES) has had limited resources to implement the necessary housing unit and person operations. These operations include developing and implementing the costly field and clerical matching procedures necessary for a sample-based application of the dual system estimation. For example, the 2010 Census coverage survey had field operation costs of $15.2 million for Independent Listing, $23.7 million for Interviewing, and $14.9 million for additional person follow-up (Contreras et al. 2012, Linse and Argarin 2012, Johnson et al. 2012). This does not include the system development costs and the 2020 PES budget estimates. The 2020 PES has a projected fieldwork budget of about $50 million.

For the 2018 End-to-End Census Test, the Post-Enumeration Survey was de-scoped, so the first time that these operations will be implemented will be during the 2020 Census. Based on these developments, we decided to propose an evaluation to examine whether administrative records could be used to produce coverage estimates similar to the survey-based results without having to implement the field collection, clerical matching software development, and clerical matching personnel costs. This evaluation will use administrative records as the second system instead of the independent PES sample. The census will continue to serve as the first system. This evaluation will begin by researching and prototyping approaches on the 2010 Census, and we will compare the results to the 2010 Census Coverage Measurement (CCM) and DA coverage estimates. Then, we will implement our approach to produce coverage estimates for the 2020 Census and again compare these with the official estimates of coverage from the 2020 PES and DA programs.

II. Background

1. Post-Enumeration Surveys for the Decennial Census

The U.S. Census Bureau has a history of producing population estimates using dual system estimation from a sample survey. Wolter (1986) gives an overview of producing a population estimate using two lists or systems and provides the assumptions that need to be met in order to produce valid population estimates. These assumptions include independence of being captured on both lists, removing erroneous inclusions from each system, having no matching error, and others.

For census applications, one system is the Enumeration or E system. This starts with the enumerations in the census. The second system is the Population or P system. This is an independent enumeration of the population. The Wolter (1986) paper outlines how dual system estimates can be produced using the Petersen model to generate valid estimates in post-strata. Synthetic estimation could then be used to produce small area estimates below the post-strata.

Hogan and Wolter (1988) document the 1980 Census coverage estimates. For the 1980 survey, the Current Population Survey was used as the second system. The 1980 analysis produced a series of alternative estimates based on different assumptions. Lessons learned from this application were incorporated into the design for the 1990 PES.

Hogan (1993) documents the production of population estimates for the 1990 PES. The 1990 PES used an independent listing of sampled census blocks instead of the Current Population Survey. Independent interviews were conducted at these addresses. This included collecting where the person should have been counted on Census Day. The PES matched the independently rostered people to the areas where they identified that they should have been counted. Computer and clerical matching were used to determine the residence, Census Day, and match statuses of the cases. Follow-up interviewers were conducted based on stringent criteria to resolve cases. The final population estimates were based on calculating the dual system estimates for 1,392 post-strata (later revised to be 357 post-strata). See Hogan (1993) for more information about the 1990 PES.

Hogan (2003) documents the production of population estimates for the 2000 Accuracy and Coverage Evaluation (A.C.E.). The 2000 A.C.E. was similar to the 1990 PES in that independent listing and interviewing on a sample basis were conducted. The 2000 A.C.E. rostered both the outmovers and inmovers since Census Day for independent sample addresses. The outmovers were matched in the sample area to get the match rate for the mover population. The inmovers were used to get an estimate of the independent mover population for a post-strata. Similar to 1990, computer, clerical, and follow-up operations were conducted.

The initial 2000 coverage survey estimates showed a difference of about 3 million people as compared with those produced by the Demographic Analysis program (U.S. Census Bureau 2001). Robinson et al. (2002) document how estimates of births, deaths, immigration, and emigration are put together to produce population estimates. Analysis of the initial A.C.E. estimates led to two sets of revisions. The final revisions accounted for erroneous inclusions in both systems and matching error. The final estimates also included an adjustment for the violation of independence between the captures (correlation bias). The main change in the estimates was due to erroneous inclusions in the census system that had not been removed for the original estimates.

Mule (2012) documents the person coverage estimates from the 2010 Census Coverage Measurement (CCM) survey. Like the 1990 and 2000 coverage surveys, the 2010 CCM evaluated net census coverage by using dual system estimation to generate population estimates. New for 2010, the CCM used logisitic regression modeling instead of post-stratification to

produce synthetic estimates of net coverage. The logistic regression modeling allowed for the reduction of correlation bias in the total population estimates without having to include unnecessary higher-order interactions as when forming post-stratification cells. Like the 2000 A.C.E., the 2010 CCM included an adjustment for remaining correlation bias for some of the population estimates.

2. Administrative Records Research for the Decennial Census

The Census Bureau has conducted research about using administrative records during the enumeration. Leggeri et al. (2002) document the administrative records census experiment in Census 2000. This was an experiment to see if administrative records could be used to conduct the decennial enumeration for two counties in Maryland and three counties in Colorado. The results showed potential undercoverage when solely using administrative records for enumeration at that time.

The 2010 Census also included multiple evaluations involving administrative records. Sheppard et al. (2013) describe an evaluation that used administrative records to detect and improve overcoverage. While not examining undercoverage, this 2010 evaluation recommended that future studies about improving census coverage with administrative records include a follow-up component to assess discrepancies between the census and administrative records. The 2010 Census Match Study assessed the quality and coverage of several administrative records and third-party sources relative to the 2010 Census (Rastogi and O'Hara 2012). Relevant to our evaluation, Rastogi and O'Hara found that the administrative records and third-party sources had less coverage for harder-to-count populations.

Keller et al. (2018) document how administrative records and third-party data are being used to reduce contacts in the Nonresponse Followup operation in the 2020 Census. This work includes building rosters from administrative record sources and determining for each address if we have enough information to reduce the number of times a fieldworker attempts to visit the address to obtain a census response. Variations of this approach were implemented in the 2013, 2014, 2015, 2016, and 2018 census tests. Results from each test were used to refine and improve the methodology for the next test. Our evaluation involves similar work of using administrative records to build rosters independent of the concurrent census.

III. Assumptions

This evaluation has the following assumptions.

1. The project team will have adequate time to implement the evaluation as it is designed in this study plan.
2. The administrative records sources, including federal tax information, that are needed to conduct this evaluation will be approved and made available to the research team. See Section VI for a list of the requested sources.
3. The administrative records sources approved for this research will be consistent between 2010 and 2020, to the extent possible. That is, the universe, format, and data contents of the administrative records sources will be the same over time. If the files are inconsistent

between 2010 and 2020, then the methods we develop on the 2010 data may not work well for 2020.

4. 2020 Census files such as the Census Unedited File (CUF) and Census Edited File (CEF) will be available for the required analysis. The 2020 CUF will be processed through the Person Identification Validation System (PVS) to apply the Person Identification Keys (PIKs) that facilitate matching between administrative records and census sources.
5. Coverage estimates for the 2020 Census will be produced via the PES and DA as in previous decennial censuses.

IV. Research Questions

Our evaluation will address the following two research questions:

1. How do the administrative record coverage estimates compare with the 2010 Census Coverage Measurement and Demographic Analysis estimates?

2. How do the administrative record coverage estimates compare with the 2020 Post-Enumeration Survey and Demographic Analysis estimates?

V. Methodology

In order for our approach to be used to generate coverage estimates for the 2020 Census, we will use the administrative records and third-party data to implement a proof of concept on the 2010 Census. This will allow our results to be compared with the 2010 Census and the official estimates of coverage. We will then implement our approach to produce coverage estimates for the 2020 Census. Based on the results of the 2010 research, we will identify the coverage estimates that we can produce. The evaluation will assess these alternative coverage results by comparing them with the 2020 Post-Enumeration Survey and the Demographic Analysis official estimates of coverage.

A. Evaluation Design

This section provides an overview of the initial methodology being considered for this project. We describe the methodology to generate estimates for the 2010 Census using census data, American Community Survey (ACS) data, administrative records, and third-party data for the 2010 time frame. The goal is to implement the approaches that we tested with the 2010 data to produce coverage estimates using the 2020 Census data.

1. Producing Administrative Records Dual System Estimates

The proposed research will see if population estimates can be generated using administrative records as the second source in dual system estimation. Our initial goal is to generate population estimates. Our analysis will see if we can generate estimates for the total population or if we are

restricted to the household population like the survey approach. Our approach will generate population estimates for different search areas. This includes tract, state, and nation. Based on having estimates for each tract, then estimates for national, site, state, county, or other geographic areas can be aggregated. While having a national search area, this approach does not require putting a response in a specific geographic location for estimation but could possibly require synthetic assumptions for generating subnational estimates.

Our secondary objective is to generate estimates of the population by age, sex, and race and Hispanic origin. For the tract level search area approach, we will attempt to generate these estimates for each tract. Again, these tract-level estimates of demographic groups could be aggregated up to produce estimates of those characteristics for the nation, states, or counties. This will be compared with results using state or national search area methods. This research will use census, administrative records, ACS, and possibly third-party information to generate these estimates.

The census will continue to be the first system for the dual system estimation. This approach will use all responses. Since we are not doing any fieldwork, no sampling of the census responses is needed. Our approach will attempt to apply the following same four criteria as used by the survey-based enumeration:

- Appropriateness
- Uniqueness
- Completeness
- Geographic correctness

For appropriateness, we will check that the person should be included in the census by using Social Security Administration information to check if the person was born after Census Day or died before Census Day. For uniqueness, we will use the Protected Identification Keys (PIKs) assigned to the census record to retain only one record for each person. Our initial rule will keep the response closer to Census Day. If the responses are on the same day, we will investigate decision rules like using the case with more item responses or other criteria. For completeness, we will have a rule about a census record having enough information to identify a single person. For this research, we will start by using a completeness rule that a PIK needs to be assigned to each record. For geographic correctness, we will research different geographic areas. This includes using tract, state, or national search areas. However, with the absence of a coverage survey, we do not have information from the survey interview to determine whether the census people were counted in the tract in which they should have been. We will begin by assuming that people are in the correct tract, and we will consider ways to address this issue.

The people from the administrative record sources will be the second system. Since we are not doing listing, clerical matching, or follow-up, no sampling of the administrative records is needed. For administrative record people, we will determine rules about which person records to include. A conservative rule for IRS 1040 responses could be to use only people from 1040 filings that were filed after census data collection started in March. Research could determine

how to use IRS 1040 deliveries in February. Our research will also determine how to use other sources depending on when they were delivered to the Census Bureau and their reference dates.

We will apply the same four criteria to the administrative record people. For completeness, we will only use person records from administrative record sources that have a PIK assigned. For appropriateness, we will confirm based on information available at that time that the administrative record person was alive on Census Day. For uniqueness, we will make sure that each administrative record person is only associated with one address. Reference date information from the different sources can be used to develop rules to associate individuals with only one address. We will implement uniqueness by only using administrative record individuals that were assigned a PIK. We will implement geographic correctness by testing different search areas to which the administrative record was assigned.

Dual system estimation requires accurate matching between the two systems. For matching, we will match based on the PIKs assigned to the census and the administrative records individuals. Based on this, we will be able to tally the number of people with PIKs who were counted in a) both the census and administrative records (N_{11}), b) only the census (N_{10}), and c) only the administrative records (N_{01}). Thus, we will have counts for three of the four cells for the traditional two-by-two table shown below.

Table 1. Dual System Estimation Example

		Administrative Records		
		In	Out	Total
Census	In	N_{11}	N_{10}	$N_{1.}$
	Out	N_{01}	N_{00}	
	Total	$N_{.1}$		N

By having information available for three of the four cells, we will research estimation approaches that produce an estimate of the size of the fourth cell. This will be different from the survey-based approach that measures the population total based on the Petersen estimator. The post-stratification approach used by the Census Bureau estimates the marginal estimate of meeting the four correctness criteria in the census and divides that by the rate of matching independent individuals to those census cases. The 2010 CCM implemented a logistic regression equivalent.

To minimize the bias because of either dependence of capture and heterogeneous capture probabilities, we will use characteristics of the people in the estimation. Since we are using only census and administrative records that have been assigned a PIK, we will have age and sex available for those cases from the Census Numident file. For race and Hispanic origin, we will consider using 2010 Census responses or the Center for Administrative Records Research and Applications (CARRA) Best Race and Hispanic origin file. Methods using race and Hispanic origin may require the development of imputation methods to assign to cases without that characteristic available.

Some initial methodologies for estimating the size of the fourth cell include the following. George and Robert (1991) provide an approach for calculating Bayes estimates for capture-recapture models. Other possibilities include log-linear modeling approaches. Cormack (1989) has an approach for using log-linear models for capture-recapture. The R package RCAPTURE implements several of these approaches. Part of this work will be attempting to identify other estimation approaches that could be used as well.

One concern with implementing the administrative record dual system estimates is that the population estimates may suggest large overcounts or undercounts. This could be because of violations of the assumptions for producing estimates. To potentially guard against this, we will see if we can use information already available about the size of the population. Our approach will be to research if the latest 5-year ACS estimate for the total population or subgroups can be used. The final population estimates could be a combination of the administrative record DSE and the ACS 5-year estimate.

Based on producing estimates of subpopulations for the tracts, these estimates can then be aggregated to produce national, state, county, or other geographic area estimates. The methodology would need to be determined to check the measures of uncertainty associated with the necessary point estimates.

If the tract-level search area approach is used then when we are aggregating up the tract estimates, this approach should not require the synthetic bias adjustment to produce root mean square error that was done in 2010. Besides potentially not having sampling error, this approach would also be able to address synthetic error that was present in the 2010 CCM estimates. The 2010 CCM estimates had synthetic error for state, county, and place estimates because a national-level fixed-effects logistic regression model was used, and the model did not have any fixed effects for these lower levels of geography. Furthermore, the 2010 CCM estimation methodology did not use small area estimation techniques like using random effects. The 2010 CCM root mean square error estimates included estimates of synthetic bias, and as a result of this additional uncertainty, none of the state, county, nor place estimates of person net coverage were statistically different from zero. This tract-level search area approach addresses the synthetic bias issue by producing population estimates at the tract-level and then aggregating those to states and other geographic areas. If our research shows that viable population estimates can only be produced from the national search area, then an estimation methodology for subnational estimates may need to be developed.

The proof of concept work on the 2010 Census data may determine that this approach of using only the census and the administrative records produces population estimates that have too much bias or have too much uncertainty. If this is happening, then we will investigate using the ACS responses collected in or around April 2010 as a third system. We can use the PIKs assigned to the ACS responses as a third set of data. This will allow estimation approaches that account for multiple systems to be investigated. The introduction of the ACS responses as a third source would require appropriate changes in the estimation methodology. The methods described so far are based on having census and administrative records available for every housing unit across the country, but the ACS is for only a sample. We would then assess if this improves the results.

2. Prototype Analysis using the 2010 Census

The first step is to develop and show how an administrative record dual system estimation approach can generate coverage estimates using 2010 data. Our research will attempt to produce net coverage estimates for the same estimation domains as in the 2010 CCM. Mule (2012) and Davis and Mulligan (2012) document the 2010 CCM person coverage results.

Evaluation Research Question #1: How do the administrative record coverage estimates compare with the 2010 Census Coverage Measurement and Demographic Analysis estimates?

We will assess our coverage results by quantifying the percent differences between the 2010 CCM estimates and our administrative record DSE results. For each estimate, we can calculate the percent difference by formula (1).

$$Percent\ Difference = \frac{2010CCM - ADRECDSE}{2010CCM} \times 100 \quad (1)$$

For state estimates where there are multiple percent differences to analyze, we will summarize the differences by using approaches used to assess population estimates. We will investigate using mean algebraic percent differences and mean absolute percent differences. Since mean absolute percent difference measures are sensitive to outliers, one alternative summary measure is to use a rescaled version developed by Coleman and Swanson (2007).

If we are successfully able to implement the tract-level search area approach, we will calculate percent undercount estimates for each tract. We will summarize the percent undercount estimates by calculating mean algebraic percent undercount estimates to assess on average how close our estimates are to the census counts. These tract analyses can be done by response rate or other tract measure groupings. This will allow us to see if this coverage approach works in areas with higher or lower responses. While we do not have official estimates to compare, we can see what the coverage results from this approach would be. We will explore calculating mean absolute percent undercounts or other measures that show the distribution of the percent undercount estimates. Formula 2 shows the percent undercount calculation.

$$Percent\ Undercount = \frac{ADRECDSE - 2010Census}{ADRECDSE} \times 100 \quad (2)$$

3. Coverage Analysis of the 2020 Census

We will implement the researched methodology to develop coverage estimates for the 2020 Census. The results of the 2010 prototype research will factor into the estimation domains that can be produced. If possible, we will see if the methodology can also be implemented on data from the 2018 End-to-End Census Test.

Evaluation Research Question #2: How do the administrative record coverage estimates compare with the 2020 Post-Enumeration Survey and Demographic Analysis estimates?

Similar to the prototype analysis of the 2010 CCM estimates, we will produce similar estimation domains as being done for the 2020 PES. We will assess our alternative coverage estimates by quantifying the percent differences between the 2020 PES and our results. We will use similar mean algebraic percent differences and mean absolute percent differences for estimation domains like states.

B. Interventions with the 2020 Census

Our analysis will require the following information and access listed below. We estimate our impact on the system resources needed for a successful 2020 Census to be very small.

- Access to the administrative records and third-party data that is stored in the Census Data Warehouse for 2010 and 2017 through 2020. These datasets have already been or are planned to be processed by the Economic Reimbursable Surveys Division (ERD). Obtaining access for this evaluation should have no impact on a successful 2020 Census.

- Access to the Protected Identification Keys assigned by ERD for the 2010 Census, 2018 End-to-End Census Test, and 2020 Census files. The 2010 assignments are already processed. The assignments for the 2018 End-to-End Census Test and the 2020 Census are planned production processes to support characteristic imputation for the census. Our evaluation simply requires access to these same datasets when they are available. Obtaining access for this evaluation should have no impact on a successful 2020 Census.

C. Implications for 2030 Census Design Decisions and Future Research and Testing

The outcome of this evaluation will provide the Census Bureau with information for the 2021 to 2025 research and testing phase and early design of the 2030 Census to determine whether this approach is a possible viable alternative to doing dual system estimation based on independent field interviews and clerical matching. If this less expensive approach is deemed a viable alternative, then additional research into the methodology could be planned for the 2021 to 2025 research and testing phase in anticipation of implementation in the 2030 Census.

VI. Data Requirements

Data File/Report	Source	Purpose	Expected Delivery Date
IRS 1040 Tax Returns TY 2008-2009, 2017, and 2019	IRS	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	TY 2008-2009 and 2017: Already available. TY 2019: Monthly starting March 2020
IRS 1099 Information Returns TY 2008-2009, 2017, and 2019	IRS	Produce administrative records dual system estimates for the 2010	TY 2008-2009 and 2017: Already available.

Data File/Report	Source	Purpose	Expected Delivery Date
		Census, 2018 Census Test, and 2020 Census.	TY 2019: Monthly starting March 2020
CMS Medicare Enrollment Database 2009-2010, 2017-2020	CMS	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2009-2010 and 2017: Already available. 2018-2020: Sept. of given year.
IHS Patient Registration 2009-2010, 2017-2020	IHS	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2009-2010 and 2017: Already available. 2018-2020: Aug. of given year.
CARRA Kidlink File	CARRA	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	Previous versions already available. Future versions near March of given year.
Census PIK crosswalk 2010, 2018, and 2020	ERD/Census Data Warehouse	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010: Already available. 2018 and 2020: Fall of given year.
Census Unedited File 2010, 2018, and 2020	ERD/Census Data Warehouse	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010: Already available. 2018 and 2020: Fall of given year.
Census Edited File 2010, 2018, and 2020	ERD/Census Data Warehouse	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010: Already available. 2018 and 2020: February of subsequent year.
Census Numident 2010, 2018, and 2020	SSA	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010 and 2018: Already available. 2020: April 2020.
American Community Survey PIK crosswalks 2010, 2018, and 2020	ERD/Census Data Warehouse	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010: Already available. 2018 and 2020: Subsequent year.
American Community Survey unswapped edited response file 2010, 2018, and 2020	ACSO	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010: Already available. 2018 and 2020: Subsequent year.
CARRA Best Race and Hispanic Origin file	CARRA	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010 and 2018: Already available. 2020: February 2020.

Data File/Report	Source	Purpose	Expected Delivery Date
2010 CCM and DA coverage estimates	DSSD (for CCM) and POP (for DA)	Compare administrative records dual system estimates to official 2010 coverage estimates	Already available
2020 PES and DA coverage estimates	DSSD (for PES) and POP (for DA)	Compare administrative records dual system estimates to official 2020 coverage estimates	PES: June 2021 DA: December 2020
Additional administrative records and third-party sources	Federal, state, and local governments. Commercial vendors.	Produce administrative records dual system estimates for the 2010 Census, 2018 Census Test, and 2020 Census.	2010 and 2018: Many files already available. 2020: Late 2019 to 2020.

VII. Risks

1. If the IRS, CMS, HIS, or other agencies do not approve the use of their data for this evaluation, then this evaluation will not be able to be completed as planned. This evaluation will use federal tax information from the IRS and patient information from CMS and IHS.

2. If the rate at which administrative records are used in the 2020 and future census is high, then the assumption of independence between the census system (which would include administrative records enumerations) and the administrative records system may not hold. The 2020 Census will use administrative records to enumerate some nonresponding households. Another evaluation project is investigating an increased use of administrative records for future censuses. Many of the administrative records sources that we plan to use for this evaluation are the same as those being used for the 2020 and future censuses.

3. If the staff need to put more time into these production activities, this evaluation may be delayed. The staff for this evaluation will be involved in production activities for the 2020 Census and 2020 PES.

4. If there is a significant increase in missing characteristic data in the 2020 Census, this may affect the ability to apply PIKs to census individuals and carry out the analysis.
One criterion to execute this study is to have sufficiently complete response data so that PIKs can be applied to census individuals. This is achieved through reporting of characteristic data.

VIII. Limitations

1. Coverage estimates from the 2010 Census are not available at low levels of geography like the census tract. The tract-level estimates we plan to develop in this evaluation will not have point of comparison.

2. While the administrative records DSEs will not be subject to sampling error, they are subject to various nonsampling errors like matching error and classification error. Unlike sampling error, nonsampling errors are difficult to quantify.

3. By only using PIKed records, we will not be able to match records for populations that cannot be assigned a PIK. These populations include people without a Social Security Number or Tax Identification Number as well as records with insufficient information for assigning a PIK. This may lead to biased coverage estimates for areas or groups with large concentrations of these populations.

IX. Issues That Need to be Resolved

None at this time

X. Division Responsibilities

Division or Office	Responsibilities
DSSD	• Develop methodology for administrative records dual system estimation. • Compile administrative records and census data for research. • Produce report(s) summarizing findings of research.
ERD/CES	• Process administrative records and census files through the PVS to assign PIKs. Make these files available to DSSD researchers.

XI. Milestone Schedule

Evaluation Milestone	Date
Conduct research and prototype analysis on the 2010 Census.	August 2018 to August 2020
Conduct coverage analysis for the 2020 Census	September 2020 to December 2021
Distribute Initial Draft Administrative Record Dual System Estimation Report to the Decennial Research Objectives and Methods (DROM) Working Group for Pre-Briefing Review	December 2021
Decennial Census Communications Office (DCCO) Staff Formally Release the FINAL Administrative Record Dual System Estimation Report in the 2020 Memorandum Series	September 2022

XII. Review/Approval Table

Role	Approval Date
Primary Author's Division Chief (or designee)	3/15/2019
Decennial Census Management Division (DCMD) ADC for Nonresponse, Evaluations, and Experiments	2/19/2019
Decennial Research Objectives and Methods (DROM) Working Group	2/19/2019
Decennial Census Communications Office (DCCO)	mm/dd/yyyy

XIII. Document Revision and Version Control History

Version/Editor	Date	Revision Description
1.0	8/31/18	Initial draft study plan for DROM review.
1.1	11/15/18	Incorporating comments from DROM review.
1.2	3/15/19	Incorporating DROM workshop comments and quality review.
1.3	4/12/19	Incorporating DCCO edits.

XIV. Glossary of Acronyms

Acronym	Definition
A.C.E.	Accuracy and Coverage Evaluation
ACS	American Community Survey
ACSO	American Community Survey Office
CARRA	Center for Administrative Records Research and Applications
CCM	Census Coverage Measurement
CEF	Census Edited File
CMS	Center for Medicare and Medicaid Services
CUF	Census Unedited File
DA	Demographic Analysis
DROM	Decennial Research Objectives and Methods
DSE	Dual System Estimates/Estimation
DSSD	Decennial Statistical Studies Division
ERD	Economic Reimbursable Surveys Division
IHS	Indian Health Service
IRS	Internal Revenue Service
PES	Post-Enumeration Survey
PIK	Person/Protected Identification Key
POP	Population Division
PVS	Person Validation System
SSA	Social Security Administration

XV. References

Coleman, C., and Swanson, D. (2007), "On MAPE-R as a measure of cross-sectional estimation and forecast accuracy" Journal of Economic and Social Measurement, Volume 32, Issue 4
https://content.iospress.com/journals/journal-of-economic-and-social-measurement/32/4

Contreras, G., Cronkite, D., Rosenberger, L., Wakim, A. and Argarin, A. (2012), "Assessment for the 2010 Census Coverage Measurement Initial Housing Unit Independent Listing, Matching and Followup Operations" 2010 Census Planning Memoranda Series No. 178 Reissue, U.S. Census Bureau.
https://www.census.gov/2010census/pdf/2010_Census_CCM_IHU_Assessment.pdf

Cormack, R.M. (1989), "Log-Linear Models for Capture-Recapture," Biometrics, June 1989
https://www.jstor.org/stable/2531485

Davis, P. and Mulligan, J. (2012) "2010 Census Coverage Measurement Report: Net Coverage for Household Population in the United States" 2010 Census Coverage Measurement Memorandum Series 2010-G-03.
https://www.census.gov/coverage_measurement/pdfs/g03.pdf

George, E. and Robert, C. (1991) "Calculating Bayes Estimates for Capture-Recapture Models" Technical Report #90-36C, University of Purdue, revised August 1991.
http://www.stat.purdue.edu/docs/research/tech-reports/1990/tr90-36c.pdf

Hogan, H. (1993), "The 1990 Post-Enumeration Survey: operations and results" Journal of the American Statistical Association, September 1993.
https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1993.10476374

Hogan, H. (2003), "The Accuracy and Coverage Evaluation: Theory and Design" Survey Methodology, December 2003.
http://www.statcan.gc.ca/pub/12-001-x/2003002/article/6782-eng.pdf

Hogan and Wolter (1988), "Measuring Accuracy in a Post-Enumeration Survey" Survey Methodology, June 1988.
http://www.statcan.gc.ca/pub/12-001-x/1988001/article/14597-eng.pdf

Johnson, S., Sanchez, P., Wakim, A. and Henderson, K. (2012), "Assessment Report for the 2010 Census Coverage Measurement Person Matching and Followup Operations" 2010 Census Planning Memoranda Series No. 242, U.S. Census Bureau.
https://www.census.gov/2010census/pdf/2010_Census_CCM_PMF_Assessment.pdf

Keller, A., Mule, V., Morris, D. and Konicki, S. (2018 upcoming), "A Distance Metric for Modeling the Quality of Administrative Records for Use in the 2020 U.S. Census," Journal of Official Statistics, accepted for publication in 2018.

Leggeri, C., Pistiner, A., and Farber, J. (2002), "Methods for Conducting an Administrative Record Experiment in Census 2000" Proceedings of the Section on Survey Research Methods. https://pdfs.semanticscholar.org/3591/cdf4581a5af83c9b2ffbad0b41a2335ca03c.pdf

Linse, K. and Argarin, A. (2012), "2010 Census Coverage Measurement Person Interview Operation Assessment" 2010 Census Planning Memoranda Series No. 208, U.S. Census Bureau. https://www.census.gov/2010census/pdf/2010_Census_CCM_PI_Assessment.pdf

Mule (2012), "2010 Census Coverage Measurement Report: Summary of Estimates of Coverage for Persons in the United States" 2010 Census Coverage Measurement Memorandum Series 2010-G-01.
https://www.census.gov/coverage_measurement/pdfs/g01.pdf

Rastogi, S. and O'Hara, A. (2012). "2010 Census Match Study Report." 2010 Census Planning Memoranda Series No. 247, U.S. Census Bureau.
https://www.census.gov/2010census/pdf/2010_Census_Match_Study_Report.pdf

Robinson, J.G., Ahmed, B., Das Gupta, P., and Woodrow, K.A. (1993). "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis" Journal of the American Statistical Association, 88: 1061-1071.
https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1993.10476375#.XLDCoGN7lpg

Robinson, J. Gregory, A. Adlakha, and K.K. West. (2002). "Coverage of Population in Census 2000: Results from Demographic Analysis" Population Research and Policy Review, April 2002.
https://link.springer.com/article/10.1023/A:1016537822148

Sheppard, D., Stewart, T., Rothhaas, C., Lestina, F., Compton, E., Machowski, J., and Smith, D. (2013). "2010 Census Administrative Records Use for Coverage Problems Evaluation Report," 2010 Census Planning Memoranda Series No. 254, U.S. Census Bureau.
https://www.census.gov/2010census/pdf/2010_census_administrative_records_use_for_coverage_problems_evaluation_report.pdf

U.S. Census Bureau (2001), "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy". March 2001.
https://www.census.gov/dmd/www/pdf/Escap2.pdf

U.S. Census Bureau (2012), "Documentation for the Revised 2010 Demographic Analysis Middle Series Estimates," U.S. Census Bureau Technical Documentation.
https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/da_methodology.pdf

Wolter (1986) "Some Coverage Error Models for Census Data" Journal of the American Statistical Association.
https://pdfs.semanticscholar.org/d5d6/ba73b10d527c6961726f23da1ebbb27b2df8.pdf