



This document was prepared by and for Census Bureau staff to aid in future research and planning, but the Census Bureau is making the document publicly available in order to share the information with as wide an audience as possible. Questions about the document should be directed to Kevin Deardorff at (301) 763-6033 or kevin.e.deardorff@census.gov

November 19, 2012

2010 CENSUS PLANNING MEMORANDA SERIES

No. 247

MEMORANDUM FOR The Distribution List

From: Burton Reist *[signed]*
Acting Chief, Decennial Management Division

Subject: 2010 Census Match Study Report

Attached is the 2010 Census Match Study Report. The Quality Process for the 2010 Census Evaluations, Experiments, and Assessments was applied to the methodology development, specifications, software development, analysis, and documentation of the analysis and results, as necessary.

If you have questions about this report, please contact Sonya Rastogi at (301) 763-6038 or Amy O'Hara at (301) 763-5757.

Attachment

2010 Census Match Study

U.S. Census Bureau standards and quality process procedures were applied throughout the creation of this report.

FINAL REPORT

Authors:

Sonya Rastogi and Amy O'Hara

Contributors:

James Noon, Ellen A. Zapata, Cindy Espinoza, Leah B. Marshall, Teresa A. Schellhamer, and J. David Brown

Center for Administrative Records Research
and Applications



This page is intentionally left blank.

Table of Contents

Executive Summary	ix
1. Introduction	1
2. Background	1
2.1 Administrative Records in Census Programs	1
2.2 Previous Household Administrative Records Research	2
3. Methodology	3
3.1 Data	3
3.1.1 Federal Data from Other Agencies	3
3.1.2 2010 Census Data	4
3.1.3 Commercial Data	4
3.1.4 Description of Data Utilized in Address, Person, and Person-Address Pairs Results Sections.....	5
3.1.5 Description of Data Utilized in Demographic Quality and Coverage Results Section... ..	5
3.2 Record Linkage	6
3.3 Count and Match Ratios	7
3.4 Best Address for Person-Address Pairs.....	8
4. Limitations	8
5. Results	9
5.1 Address Count and Match	9
5.2 Person Count and Match	24
5.3 Person-Address Pair Count and Match	37
5.4 Demographic Quality and Coverage Assessment	50
6. Related Census Program for Evaluations and Experiments Reports	67
7. Lessons Learned, Conclusions, and Research Implications	67
8. Acknowledgements	72

Appendix 1. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by State	75
Appendix 2. Number of Administrative Records Race Response Data that Matched to the 2010 Census	76
Appendix 3. Number of Administrative Records Age Response Data that Matched to the 2010 Census	77
Appendix 4. Number Coverage of 2010 Census Race Data by Administrative Records Source Files	78
Appendix 5. Number Coverage of 2010 Census Age Data by Administrative Records Source Files	79

List of Tables

Table 1. 2010 Census and Administrative Records Address Count and Match	11
Table 2. 2010 Census and Administrative Records Address Count Ratio, Match Ratio, and Type of Enumeration Area for the Ten States with the Lowest and Highest Ratios	13
Table 3. 2010 Census and Administrative Records Address Count Ratio, Match Ratio, and Type of Enumeration Area for the Ten Counties with the Lowest and Highest Ratios.....	16
Table 4. 2010 Census and Federal and Commercial Administrative Records Address Count and Match Numbers and Ratios.....	19
Table 5. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by Type of Enumeration Area	20
Table 6. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by Housing Unit Type	21
Table 7. 2010 Census and Administrative Records Address Match by Race and Hispanic Origin of Householder, Mode, Imputation, and Proxy.....	23
Table 8. 2010 Census and Administrative Records Person Count and Match Numbers and Ratios by Region	27
Table 9. 2010 Census and Administrative Records Person Count and Match Numbers and Ratios by State.....	28
Table 10. 2010 Census and Federal and Commercial Administrative Records Person Count and Match Numbers and Ratios.....	32
Table 11. 2010 Census and Administrative Records Person Count and Match Numbers and Ratios by Type of Enumeration Area	33
Table 12. 2010 Census and Administrative Records Person Match by Demographic Characteristics, Mode, and Proxy	34
Table 13. 2010 Census and Administrative Records Person-Address Count and Match Numbers and Ratios by Region	39
Table 14. 2010 Census and Administrative Records Person-Address Count and Match Numbers and Ratios by State.....	40
Table 15. 2010 Census and Administrative Records Person-Address Count Ratio, Match Ratio, and Type of Enumeration Area for the Ten Counties with the Lowest and Highest Ratios.....	43

Table 16. 2010 Census and Federal and Commercial Administrative Records Person-Address Count and Match Numbers and Ratios	44
Table 17. 2010 Census and Administrative Records Person-Address Count and Match Numbers and Ratios by Type of Enumeration Area.....	45
Table 18. 2010 Census and Administrative Records Person-Address Match by Race, Hispanic Origin, Age, Sex, Mode, and Proxy.....	46
Table 19. 2010 Census and Administrative Records by Housing Unit Status.....	48
Table 20. 2010 Census and Administrative Records Housing Unit Status by Mode	49
Table 21. 2010 Census and Administrative Records Population Count at an Address	49
Table 22. Difference in Population Count, when Administrative Records had a Higher or Lower Population Count Relative to the 2010 Census.....	50
Table 23. Number and Percentage of Administrative Records Hispanic Origin Response Data that Matched to the 2010 Census	51
Table 24. Percentage of Administrative Records Race Response Data that Matched to the 2010 Census	52
Table 25. Percentage of Administrative Records Age Response Data that Matched to the 2010 Census	55
Table 26. Number and Percentage of Administrative Records Sex Response Data that Matched to the 2010 Census	56
Table 27. Coverage of 2010 Census Demographic Data by Administrative Records Demographic Response Data.....	57
Table 28. Coverage of 2010 Hispanic Origin Data by Administrative Records Hispanic Origin Response Data by Mode	58
Table 29. Coverage of 2010 Race Data by Administrative Records Race Response Data by Mode	59
Table 30. Coverage of 2010 Age Data by Administrative Records Age Response Data by Mode	61
Table 31. Coverage of 2010 Sex Data by Administrative Records Sex Response Data by Mode	62

Table 32. Number and Percent Coverage of 2010 Hispanic Origin Data by Administrative Records Source Files.....	63
Table 33. Percent Coverage of 2010 Race Data by Administrative Records Source Files	64
Table 34. Percent Coverage of 2010 Age Data by Administrative Records Source Files.....	65
Table 35. Number and Percent Coverage of 2010 Sex Data by Administrative Records Source Files.....	66

List of Figures

Figure 1. Count and Match of 2010 Census and Administrative Records Addresses	10
Figure 2. Count Ratio of 2010 Census and Administrative Records Addresses by County.....	15
Figure 3. Match Ratio of 2010 Census and Administrative Records Addresses by County.....	17
Figure 4. Count and Match of 2010 Census and Administrative Records Persons.....	25
Figure 5. Count Ratio of 2010 Census and Administrative Records Persons by County.....	30
Figure 6. Match Ratio of 2010 Census and Administrative Records Persons by County.....	31
Figure 7. Count and Match of 2010 Census and Administrative Records Person-Address Pairs.....	38
Figure 8. Count Ratio of 2010 Census and Administrative Records Person-Address Pairs by County.....	41
Figure 9. Match Ratio of 2010 Census and Administrative Records Person-Address Pairs by County.....	42

Executive Summary

Study Overview

To reduce costs many countries use administrative data to assist in censuses or as a replacement to traditional censuses (Farber and Leggieri 2002, Ralphs and Tutton 2011). Currently administrative data are utilized in numerous, critical U.S. Census Bureau programs for population, economic, income and poverty, and health insurance estimates, but administrative data have not yet been extensively used to assist in decennial census operations. The Census Bureau is researching ways in which to use administrative data in decennial census operations to reduce costs.¹ This study, building and expanding on previous research that utilized Census 2000 results, provides a foundation for decennial census operational research on administrative records by assessing the quality and coverage of administrative data relative to the 2010 Census.

In the United States, decennial censuses determine apportionment of state representation to Congress, are used in state redistricting, and are used to distribute billions of federal dollars (Reamer 2010). While households are required by law to participate in the decennial census, there are many households that do not respond to initial contact attempts. This requires the Census Bureau to send enumerators door to door to collect data from non-responding households in decennial census operations called Nonresponse Followup Operations.² This effort is expensive for the Census Bureau and was estimated to cost around 1.4 billion dollars in Census 2000 of a total census budget of six billion dollars (Farber and Leggieri 2002, Walker et al. 2012). The estimated cost of these operations in the 2010 Census was about two billion dollars (Walker et al. 2012). Administrative records may be able to assist with expensive operations such as Nonresponse Followup Operations, which would save the government and taxpayers a substantial amount of money.

Census Bureau staff conducted research on the use of federal administrative data utilizing Census 2000 results. The Statistical Administrative Records System (StARS) was developed from select federal data sources in 1999. Decennial census research using these data included address and person count comparisons relative to Census 2000 (Farber and Leggieri 2002). StARS 1999 was also utilized in a field test that simulated a census in several counties during Census 2000 (Berning 2003, Bye and Judson 2004).

The 2010 Census Match Study builds on this research by evaluating the federal data sources used in StARS, additional federal data sources, and commercial data. This report is also distinctive from past research in that it matches addresses and persons in administrative records to the 2010 Census to evaluate the quality and coverage of administrative data. The matching is conducted

¹ For the purposes of this report, “administrative data” and “administrative records” are used interchangeably.

² Nonresponse Followup Operations include Nonresponse Followup, Nonresponse Followup Reinterview, Nonresponse Followup Vacant Delete Check, and Nonresponse Followup Residual. For more information, see Walker et al. (2012).

using unique address and person identifiers called master address file identification numbers and protected identification keys assigned by the Person Identification Validation System to addresses and persons in the 2010 Census and administrative records. Using count and match ratios, this study evaluates the administrative data and the 2010 Census at different levels of geography and by factors such as Hispanic origin, race, and mode of data collection. This report also evaluates the quality and coverage of Hispanic origin, race, sex, and age response data in administrative records relative to the 2010 Census.

Results Overview

Addresses

There were 131.7 million addresses in the 2010 Census and 151.3 million addresses in administrative records. Of the 2010 Census addresses, administrative records matched to 122.0 million or 92.6 percent; 29.3 million administrative records addresses were not found in the 2010 Census; and 9.7 million addresses were in the 2010 Census, but not in administrative records.

Definitional differences between addresses in the 2010 Census and administrative records contributed to the address non-matches. For instance, there were Post Office Box addresses in administrative data but none in the 2010 Census. The 2010 Census also contained physical descriptions for addresses such as “yellow house near fork in the road” that cannot be matched to administrative records. Additionally, administrative records contained non-residential addresses and may have contained new construction that was not recorded in the 2010 Census.

Persons

The person match ratios were lower than the match ratios for addresses. This is in part because all addresses in the 2010 Census had master address file identification numbers, thus all 2010 Census addresses had the potential to be matched to administrative records addresses with master address file identification numbers. However, in the 2010 Census, not all persons received a protected identification key, reducing the number of persons in the 2010 Census that had the potential to match to administrative records. Protected identification keys were assigned through probabilistic matching to records using name, address, and date of birth information.

There were 308.7 million persons in the 2010 Census, and 279.2 million were assigned a protected identification key. There were 312.2 million unique persons in administrative records that were assigned a protected identification key and were alive on Census Day, April 1, 2010. Administrative records matched to the vast majority of persons in the 2010 Census that received a protected identification key, 273.6 million or 98.0 percent. The percentage of the entire 2010 Census universe, including records lacking protected identification keys, with matching administrative records was lower at 88.6 percent.

There were 29.6 million 2010 Census persons that did not receive a protected identification key. There were 48.8 million administrative records that were assigned a protected identification key, but did not match to the 2010 Census. Future research will study the potential overlap between these universes.

There were 5.5 million 2010 Census persons with protected identification keys that were not found in administrative records data, and most of them were under the age of 17. There were several reasons why administrative data did not cover children as well as other age groups, including timing issues with tax data. Tax return data from the previous tax year failed to include babies born after January 2010, however these children would likely be reported in the 2010 Census, resulting in a lower match between administrative records and the 2010 Census for babies.

Person-Address Pairs

The match ratios for person-address pairs (i.e. a person at an address) were lower relative to the address results and person results, in part because the person-address pair data incorporate both address and person matching issues, including the presence of multiple addresses for persons in administrative records. Of the 312.2 million persons in administrative records that had a protected identification key, 301.5 million had a master address file identification number and 10.7 million did not have a master address file identification number. There were 216.2 million person-address pairs in the 2010 Census that matched to administrative records. Of the 308.7 million persons-address pairs in the 2010 Census, 70.0 percent matched to administrative records person-address pairs. Of the 279.2 million person-address pairs in the 2010 Census that had a protected identification key, 77.4 percent matched to administrative records person-address pairs.

After the best address model was applied to persons in administrative records with multiple addresses in administrative records, there were 203.2 million person-address pairs in the 2010 Census that matched to administrative records. Of the 308.7 million persons in the 2010 Census, 65.8 percent matched to administrative records person-address pairs. Of the 279.2 million person-address pairs in the 2010 Census that had a protected identification, 72.8 percent matched to administrative records person-address pairs. There were 98.6 million administrative records person-address pairs that did not match to the 2010 Census. There were 76.0 million person-address pairs that were in the 2010 Census which did not match to person-address pairs in administrative records.

Demographic Quality and Coverage

The quality of Hispanic origin response data from federal and commercial files, as defined by response match ratios between the 2010 Census and administrative data, ranged from 29.4 percent to 93.1 percent. Overall, federal data sources tended to have higher quality race data for each race group relative to the commercial data. The quality of race data varied by race group.

The White alone, Black alone, and Asian alone populations tended to have higher quality race data in administrative records compared to Two or More Races, Native Hawaiian or Other Pacific Islander alone, American Indian or Alaska Native alone, and Some Other Race alone populations.

Federal and commercial files had high quality data for age and sex responses. Across federal and commercial files that had date of birth information, the age match ratio ranged from 79.0 percent to 98.5 percent. The sex match ratios ranged from 94.7 percent to 100.0 percent.

The demographic coverage analysis evaluated whether administrative data provided a demographic response to Hispanic origin, race, age, and sex groups in the 2010 Census regardless of the quality of the response. There was a Hispanic origin response present in administrative data for 92.2 percent of non-Hispanic respondents and 78.9 percent of Hispanics in the 2010 Census. The race response coverage in administrative records ranged from 46.1 percent for the Some Other Race alone population to 81.0 percent for the White alone population. Coverage by age group ranged from 84.9 percent to 94.3 percent with older age groups achieving higher coverage relative to younger age groups. Coverage for sex was 90.1 percent, where females had slightly higher coverage (90.8 percent) relative to males (89.3 percent).

Research Implications

1. **Administrative records can enhance, but not replace the decennial census.** While the quality and coverage of administrative records relative to the 2010 Census suggests that administrative records can be utilized in decennial census operations, the quality is not high enough and the coverage is not expansive enough to replace a traditional census.
2. **Use of administrative records in Nonresponse Followup can reduce costs.** Administrative records cover a substantial number of Nonresponse Followup addresses and persons, and nearly half of person-address pairs. Of the 23.6 million addresses that responded in Nonresponse Followup in the 2010 Census, administrative records matched to 21.0 million or 89.2 percent.³ Administrative records also matched to a substantial number of persons that were in Nonresponse Followup in the 2010 Census. Of the 60.4 million persons in Nonresponse Followup in the 2010 Census, 48.0 million or 79.5 percent were in administrative records. Administrative records matched to a lower number and proportion of person-address pairs in Nonresponse Followup compared to addresses and persons. Of the 60.4 million 2010 person-address pairs in Nonresponse Followup, there were 28.7 million or 47.5 percent that matched to administrative records.

³ There are 47.2 million housing units in Nonresponse Followup according to the “2010 Census Nonresponse Followup Operations Assessment” (see Walker et al. (2012)). This number is much higher relative to the housing units in this report for several reasons. For instance, the number of Nonresponse Followup housing units in Walker et al. (2012) include vacant, deletes, and unresolved households, whereas the Nonresponse Followup housing units in this report are all occupied.

Research and improvements in record linkage, refinements of the best address model, and acquiring data that cover those most likely to be in Nonresponse Followup may enhance the person-address match between the 2010 Census and administrative records.

3. **Administrative records can assist in determining housing unit and occupancy status.**

Administrative records can assist to verify whether a housing unit is a valid livable housing unit and whether it is occupied. Occupancy status results demonstrate the value of administrative records for these purposes. Of the 116.7 million occupied housing units in the 2010 Census, administrative records indicated that 96.1 million or 82.3 percent were occupied. The 2010 Census designated 15.0 million housing units as vacant, of which administrative records found that 11.4 million or 76.1 percent were not occupied. Of the 4.9 million housing units designated as deletes in the 2010 Census, administrative records indicated that 4.2 million or 85.4 percent were not occupied.⁴

4. **Administrative records can inform household population count assignment.**

Administrative records had the same population count for the majority of 2010 Census housing units that matched to administrative records. Of the 116.7 million 2010 Census occupied housing units, 96.1 million matched to administrative records. Of these, 55.5 million or 57.7 percent of housing units had the same population count. When administrative records and the 2010 Census did not have the same population count, the count differed by one person for 63.7 percent of the housing units. Further research should be conducted on this universe.

5. **Acquiring additional federal, state, and commercial data can improve address, person, and demographic characteristic coverage.** Administrative data do not cover children as well as they cover adults. Also, the quality of race and Hispanic origin response data from federal and commercial sources varies considerably by race and Hispanic origin group. The Census Bureau should partner with federal agencies, state agencies, community groups, and other organizations to obtain data that contain information on children living in households, and additional race and Hispanic origin response data should be acquired, particularly for groups where the quality of race or Hispanic origin response data is low in administrative records. Obtaining data for the following groups should be a priority: Two or More Races, Native Hawaiian or Other Pacific Islander, and American Indian or Alaska Native.

6. **Administrative records can inform race and Hispanic origin determination.** For some race and Hispanic origin groups, the quality of administrative records response data was high. For instance, the White alone, Black alone, and Asian alone populations had

⁴ Deletes refer to housing units designated for deletion from the address list. Housing units may be identified as deletes for a number of reasons including being demolished, uninhabitable, or nonresidential. Counts of 2010 Census addresses designated as deletes may vary across 2010 Census Program for Evaluations and Experiments reports as a result of different data sets being used for analysis.

relatively high quality race response data in administrative records compared to other race groups. The quality of administrative records files ranged from 94.7 percent to 99.1 percent for the White alone population. The quality of federal data for the Black alone population ranged from 87.4 percent to 98.3 percent. The range was considerably lower for commercial data. For the Asian alone population, the quality of both federal and commercial data ranged from 58.0 percent to 94.1 percent. Data could also be used for other race groups from administrative records, but the quality was generally lower. Research should be conducted on how administrative records can assist with race and Hispanic origin determination for censuses and surveys.

7. **Administrative records can assist age and sex determination.** The quality of age and sex response data in administrative records is high. For sex, the quality of administrative data ranged from 94.7 percent to 100.0 percent across administrative records files. For age, in data sources that contained date of birth, the quality of administrative records ranged from 79.0 percent to 98.5 percent. Research should be conducted on how administrative data can assist with age and sex determination for censuses and surveys.
8. **Conduct additional record linkage research with the aim of improving match results for unvalidated person records.** Many improvements were made to the Person Identification Validation System to enhance the assignment of protected identification keys and master address file identification numbers to administrative records data. Continued record linkage research on the Person Identification Validation System should be conducted to further enhance the assignment of protected identification keys and master address file identification numbers to persons and addresses, potentially increasing the universe of persons and addresses that can be matched and unduplicated between censuses and surveys and administrative records. For instance, of the 308.7 million persons in the 2010 Census, 29.6 million did not receive a protected identification key. Of these, 10.3 million could not be sent through Person Identification Validation System processing because they lacked name and date of birth, and 19.3 million went through Person Identification Validation System processing but failed to receive a protected identification key. Additional research should be conducted on how to minimize this latter universe.
9. **Conduct record linkage research to improve match results for records with incomplete name and date of birth data.** Commercial data sources often lack complete name and date of birth information. Research to unduplicate these records that failed the Person Identification Validation System, and assess the quality of the data is needed. Research on how to use records that lack personally identifiable information is needed, moving the matching approach beyond validation using the Social Security Administration Numerical Identification File.

10. Conduct record linkage research that improves person record unduplication.

Current record linkage techniques must determine whether two people that look similar are indeed the same person or if they are two different people. Refinements on record linkage techniques will help to more accurately unduplicate person records.

11. Develop partnerships with federal and state agencies to better understand administrative records and enhance record linkage research. Partnering with federal and state agencies will facilitate knowledge sharing on the availability of data that could enhance record linkage processes. This knowledge sharing will also benefit administrative records research. For instance, a better understanding of how data were collected could assist in the validation and unduplication process and improve understanding of resulting linkages.

12. Assess whether an administrative records composite improves missing data assignment. Building an administrative records composite involves unduplicating records, assigning persons at multiple addresses to one address, and assigning one characteristic to people that have different characteristics across source files. Research should assess the quality of missing data assignment using a composite compared to using all available administrative data.

13. Analyze linked survey data, especially the American Community Survey, to explore characteristics associated with data coverage and consistency. Evaluating administrative records relative to the 2010 Census provided important information, at different levels of geography and by certain characteristics, about the quality and coverage of administrative data. Other evaluations using survey data such as the American Community Survey can provide additional insights because the American Community Survey has many additional characteristics that can be analyzed.

1. Introduction

Countries are increasingly adopting the use of administrative records within surveys and censuses to reduce costs. Many European countries such as Austria, Denmark, Finland, the Netherlands, Sweden, and Switzerland already use administrative records in part or entirely in their censuses (Farber and Leggieri 2002, Ralphs and Tutton 2011). Other countries such as England, Canada, Israel, and Italy are researching ways in which to use administrative records in their censuses (Ralphs and Tutton 2011).

At the Census Bureau, uses of administrative records have expanded over the years and are critical to the success of many programs including the Business Register, Intercensal Population Estimates, Local Employer Dynamics, Demographic Analysis Estimates, Small Area Income and Poverty Estimates, and Small Area Health Insurance Estimates. However, the use of administrative records has not been widely adopted within decennial census operations.

External researchers in the 1980s and the National Academies of Sciences in the mid-1990s called for research to be undertaken on the use of administrative records in decennial census operations (Alvey and Scheuren 1982, Edmonston and Schultze 1995, Steffey and Bradburn 1994). This spurred the Census Bureau to develop StARS in 1999. StARS 1999 was constructed and evaluated by Census Bureau staff, and utilized in a Census 2000 field test that simulated an administrative records census in several counties (Farber and Leggieri 2002, Berning 2003, Bye and Judson 2004). The 2010 Census Match Study builds upon and expands this research.

The 2010 Census Match Study is the first study that links administrative records to decennial census results to evaluate the quality and coverage of administrative records. This study evaluates counts and matches of addresses and persons, and persons at addresses at different levels of geography and by factors such as Hispanic origin, race, and mode of data collection. This report also evaluates the quality and coverage of Hispanic origin, race, sex, and age data in administrative records relative to the 2010 Census.

2. Background

2.1 Administrative Records in Census Programs

Many important programs at the Census Bureau utilize administrative records extensively. Administrative records are used to update the Business Register, the survey frame for the Economic Census, and most monthly, quarterly, and annual economic surveys. The Population Estimates program utilizes administrative birth and death data, as well as data from Medicare, to produce annual estimates of the U.S. population at the national, state, and county levels. Uses of these estimates include federal funding allocations and survey controls. Additionally, the Local Employment Dynamics program utilizes labor market data from states to develop critical information on employment, job creation, turnover, and earnings. Demographic Analysis

Estimates utilize administrative birth and death data, as well as data from Medicare, to assess the coverage of decennial censuses.

To help inform the administration of federal programs and the allocation of federal funds to local jurisdictions, the Small Area Income and Poverty Estimates program develops current selected income and poverty estimates for states, counties, and school districts using a combination of American Community Survey (ACS) data, administrative records, population estimates, and decennial census data. The Small Area Health Insurance Estimates program provides health insurance coverage estimates for states and counties from statistical model-based methods using survey, decennial census, and administrative data sources.

While administrative data have been incorporated into a number of important Census Bureau programs, it has not yet been highly utilized in decennial census operations. Research conducted utilizing Census 2000 results, this report, other 2010 Census Program for Evaluations and Experiments reports, and 2020 Census research will help determine the feasibility of using administrative data in decennial census operations.

2.2 Previous Household Administrative Records Research

In response to calls from external researchers and the National Academies of Science, the Census Bureau developed StARS 1999 to research the use of administrative data in decennial census operations. StARS 1999 was assembled from six administrative records sources: (1) Internal Revenue Service (IRS) Individual Income Returns, (2) IRS Information Returns, (3) Department of Housing and Urban Development (HUD) Tenant Rental Assistance Certification System (TRACS), (4) Center for Medicare and Medicaid Services (CMS) Medicare Enrollment Database (MEDB), (5) Indian Health Service (IHS) Patient Registration System, and (6) Selective Service System (SSS) Registration System (Farber and Leggieri 2002). In StARS 2000, and for subsequent years, an additional source file was added, (7) the HUD Public and Indian Housing Information Center (PIC) file.

The StARS 1999 data were assembled to test the feasibility of acquiring, validating, and unduplicating federal administrative data. The resulting files were primarily used for count comparisons relative to Census 2000 and in a Census 2000 field test called the Administrative Records Census Experiment or AREX 2000. StARS 1999 research found that address and person counts in StARS were relatively close to the counts in Census 2000 at the national level. StARS 1999 also produced counts that were similar to Census 2000 in states in the Midwest and Northeast, but there were more discrepancies with counts in the South and Southwest. Farber and Leggieri (2002) concluded that more research needed to be conducted to produce better race and ethnicity counts.

AREX 2000 investigated the possibilities of conducting an administrative records census and of using administrative records in support of a traditional census (Berning 2003). Census 2000 results for two Maryland and three Colorado counties were compared to administrative data from

StARS 1999. Nearly a one-year lag existed between the reference period of Census 2000 and several of the administrative data sources.

Count coverage of administrative data across the test counties varied according to the methodology that was used. The study also identified fewer children and more elderly people than Census 2000. Difficulties were also identified in determining the correct residence for movers. The lag between the various administrative records data reference periods and Census Day, April 1, 2000, likely contributed to these difficulties (Bye and Judson 2003).

The research on StARS 1999 and AREX 2000 provided important insights regarding the use of administrative records for decennial census operations. The 2010 Census Match Study extends the administrative records research by utilizing four additional federal files and nine commercial datasets, in addition to the data used to construct StARS. The 2010 Census Match Study also utilizes data that were close to an April 1, 2010 reference date.

3. Methodology

3.1 Data

The following sections briefly describe the federal and commercial data that were utilized in this report.

3.1.1 Federal Data from Other Agencies

Two files were used from the IRS, the Individual Income Tax Returns 1040 and Information Returns 1099. Individual Income Tax Returns provide data for individuals who file a 1040 tax return. These data include all returns received by the IRS and include the mailing address on the return (generally as of around April 15, 2010), the name and Taxpayer Identification Number (TIN) for the primary filer, and the name and TIN for any spouse and/or up to four dependents on the form. Information Returns 1099 include name, address, and TIN for individuals as reported to the IRS by financial institutions and employers on the various Information Returns (1099 forms, W2 forms, etc.).

Three files were used from HUD. The PIC data are maintained by HUD for persons participating in the public housing program and other rental assistance programs. TRACS contains data for persons receiving rental assistance and participating in other assisted housing programs through HUD. Computerized Homes Underwriting Management System (CHUMS) contains data for persons who have obtained or applied for mortgages insured under HUD/Federal Housing Administration mortgage insurance programs. These files include information such as name, address, date of birth or age, sex, race, Hispanic origin, and Social Security Number (SSN).

The 2010 Social Security Administration (SSA) Supplemental Security Record (SSR) file includes address, personal identifiers, and date of birth for Supplementary Security Income (SSI) recipients. The 2010 Census Match Study primarily used 2010 SSR files for SSI recipients and appended information on children and spouses from a separate 2011 SSR file.

The MEDB from the CMS contains Medicare enrollee data and name, address, date of birth, race, Hispanic origin, sex, and SSN. The SSS Registration File contains address and date of birth information on males, ages 18 to 25, who register with Selective Services for the purpose of creating a database which would be used in the event of a draft.

The IHS Patient Registration File contains information on American Indians or Alaska Natives (AIAN) who participate in the IHS System. Spouses and children of AIANs that are not in this race group are eligible to receive these services as well.

The National Change of Address file is maintained by the U.S. Postal Service and includes name, address, and move information such as the move date, the original address, and the new address.

Temporary Assistance for Needy Families (TANF) files include national level data for adults and children who participate or receive benefits through states' TANF programs. These files include SSN, date of birth, sex, race, Hispanic origin, and basic geographic information (state, county, and zip code). Since addresses were not included in this file, TANF is only used for the person and demographic quality and coverage sections of this report.

The Death Master File from SSA was not used in the quality and coverage analysis of administrative records relative to the 2010 Census, but assisted in processing the administrative files. It contains date of death and SSN for deaths that have been reported to SSA. Date of death information was used to help determine whether a person in administrative records was alive as of April 1, 2010.

3.1.2 2010 Census Data

The Census Edited File was used for this report. This file includes the same address and person data from the Census Unedited File along with edited demographic variables and edit and imputation flags.

3.1.3 Commercial Data

Nine data files containing identifying information and demographic characteristics were acquired from five commercial data vendors for the 2010 Census Match Study evaluation. These data are described below.⁵

⁵ Commercial data vendors are described by name in the Methodology section of this report, but all results in the Address, Person, and Person-Address sections reflect aggregated and unduplicated commercial data. License agreements with each vendor prohibit direct comparisons across companies. In the Demographic Quality and Coverage Assessment section, information about individual vendors is presented but vendor names are withheld.

The Census Bureau obtained multiple datasets from three vendors, Experian, Targus, and the Veteran Service Group of Illinois (VSGI). The Experian In-Source (INS) file contains current address, name, race, Hispanic origin, age, and sex data from credit bureau header information. The Experian End-Dated Records (EDR) file is a historical file that contains the same variables as Experian INS. The Targus Federal Consumer file contains address, name, race, Hispanic origin, age, and sex data. The Targus Pure Wireless file contains name, age, sex, and some address data. The Targus National Address File (NAF) contains addresses.

The VSGI Name and Address Resource Consumer (NAR) file contains current address, name, date of birth, race, Hispanic origin, and sex information from magazine/periodical change of address information, utility records, and other sources. The VSGI TrackerPlus (TRK) file is a historical file that contains the same variables as VSGI NAR. The VSGI race and Hispanic origin data were not used in this report, as they were at the tract level rather than at the individual level, thus quality and coverage of individual race and Hispanic origin data could not be assessed from this data source.

The InfoUSA file contains current and historical address, name, race, Hispanic origin, age, and sex data from sources such as property taxes, voter registration rolls, and telephone book white pages. The Melissa Data Base Source (Melissa) file contains address, name, and age information from credit header records, utility bills, cellular phone records, and the U.S. Postal Service.

3.1.4 Description of Data Utilized in Address, Person, and Person-Address Pairs Results Sections

All the federal and commercial data except for TANF and the Targus NAF were used in the address, person, and person-address pair result sections of this report. TANF data could not be used for the address or person-address pair evaluation as TANF did not include addresses on the file. The Targus NAF was not used for the person and person-address pair sections as the file does not contain person data.

3.1.5 Description of Data Utilized in Demographic Quality and Coverage Results Section

The demographic quality and coverage analysis used select files that contained race, Hispanic origin, age, and sex data. For race, all three HUD files, IHS, MEDB, TANF, Experian EDR, Experian INS, InfoUSA, and Targus Federal Consumer were used. The Hispanic origin analysis used all of the same files that were used for race except IHS. The sex analysis included all the files used in the race analysis plus SSS, Targus Wireless, VSGI NAR, and VSGI TRK. The age analysis included all the same files as the sex analysis plus SSR and Melissa data. In addition to these files, for all demographics, previous census records (Census 2000 and ACS 2001 to ACS 2009) and the SSA Numerical Identification File (Numident) were also evaluated. The Numident includes SSN, name, date of birth, sex, and race data for all persons who have been

assigned a SSN by the SSA. It does not include address or location information associated with records on the file, and as such it was not used in the address, person, or person-address pair results sections of this report.

The federal and commercial data do not uniformly collect and report data on Hispanic origin and race. Regarding the Numident, the SSA collected race data from 1936 to 1980 via the Social Security application based on the three categories of “White,” “Black,” and “Other.” In 1980, SSA changed its categories to “White,” “Black,” “Hispanic,” “Asian, Asian American, or Pacific Islander,” and “American Indian or Alaskan Native” in order to comply with the 1977 Office of Management and Budget (OMB) Directive 15 on Race and Ethnic Standards. The SSA then halted collecting race data when it transitioned to the Enumeration at Birth system in 1987.

The remaining federal files report race according to the OMB revised 1997 race and ethnic standards.⁶ However, unlike the Census Bureau, HUD CHUMS, HUD PIC, and TANF do not include a category for Some Other Race (SOR). While it does include this category, MEDB models its race data and does not include a category for Two or More Races.⁷ IHS differs in that it only identifies persons as AIAN and non-AIAN. The commercial files model race data and do not model more than one race for an individual.

3.2 Record Linkage

The same people and addresses are present in many of the same administrative records data sources. The administrative records files must be unduplicated in order to evaluate them relative

⁶ When collecting and tabulating data on race and ethnicity, federal agencies must adhere to guidance from the U.S. Office of Management and Budget’s (OMB) 1997 *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. The standards are available online at www.whitehouse.gov/omb/fedreg/1997standards.html.

OMB requires federal agencies to use a minimum of two ethnicities: Hispanic or Latino and Not Hispanic or Latino. Hispanic origin can be viewed as the heritage, nationality group, lineage, or country of birth of the person or the person’s parents or ancestors before their arrival in the United States. People who identify their origin as Hispanic, Latino, or Spanish may be any race. “Hispanic or Latino” refers to a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race.

OMB requires federal agencies to use a minimum of five race categories: White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander. For respondents unable to identify with any of these five race categories, OMB approved the Census Bureau’s inclusion of a sixth category, Some Other Race. The 1997 standards require federal agencies to permit respondents to self-identify with more than one race. For more information on how race was collected and tabulated in the 2010 Census please refer to Humes, K., N. Jones, and R. Ramirez. 2011. *Overview of Race and Hispanic Origin: 2010*, U.S. Census Bureau, 2010 Census Briefs, C2010BR-02, available at www.census.gov/prod/cen2010/briefs/c2010br-02.pdf.

⁷ Individuals who responded to the question on race by indicating only one race are referred to as the *race-alone* population or the group that reported only one race category. Six categories make up this population: White alone, Black or African American alone, American Indian or Alaska Native alone, Asian alone, Native Hawaiian or Other Pacific Islander alone, and Some Other Race alone. Individuals who chose more than one of the six race categories are referred to as the Two or More Races population. All respondents who indicated more than one race can be collapsed into the Two or More Races category which, combined with the six race-alone categories, yields seven mutually exclusive and exhaustive categories. Thus, the six *race-alone* categories and the Two or More Races category sum to the total population.

to the 2010 Census.⁸ Thus, unique address identifiers called master address file identification numbers (MAFIDs) and person identifiers called protected identification keys (PIKs) were assigned to administrative records through the Person Identification Validation System (PVS). To match administrative records data to the 2010 Census, MAFIDs and PIKs must be on these data sources. The 2010 Census data already had MAFIDs, therefore only PIKs were assigned to the 2010 Census through PVS. For more information on this record linkage system see Wagner and Layne (2012).

The process of assigning address identifiers starts with matching administrative data to an extract from the Census Bureau Master Address File (MAF).⁹ MAFIDs were assigned to administrative records with address data that matched to the MAF. The process of assigning PIKs to the 2010 Census and administrative data starts with matching these data to a reference file containing data on individuals.

For the assignment of PIKs, the matching software compared personally identifiable information (PII) from administrative data and the 2010 Census to PII on person reference files. The software has two primary components, and one or both of those components can be utilized depending on the characteristics available in the administrative records and 2010 Census files. The two components are “verification” and “search.” The verification module was used when the source file contained a SSN.¹⁰

Many federal administrative files contained SSNs, but the 2010 Census and most commercial data did not include SSNs. For these data sets, the search modules in the software compared name, address, and date of birth fields to the person reference file. Administrative and 2010 Census records that matched to the person reference file through either the “verification” or “search” modules were considered validated and were assigned a PIK.

3.3 Count and Match Ratios

Count and match ratios are used to evaluate the quality and coverage of administrative data relative to the 2010 Census. The count ratio is calculated by dividing the unduplicated administrative records count by the 2010 Census count and multiplying the result by 100. When the administrative records data have the same proportion of addresses, persons, or person-address

⁸ The 2010 Census also contains duplicates. Preliminary research that unduplicated the 2010 Census by PIK suggests there were 10.5 million duplicates in the 2010 Census. This is close to the official Census Coverage Measurement figures which suggest there were 8.5 million duplicates in the 2010 Census (Mule 2012). The 2010 Census Match Study report only uses the unduplicated 2010 Census for one analysis, the demographics quality analysis. Duplicates may vary by demographic group, potentially inflating quality of data for some groups while deflating quality for others. Thus, 2010 Census duplicate PIKs were removed from the demographics quality analysis.

⁹ The 2010 Census Match Study uses a Master Address File extract. For the purposes of the report, this will be referred to as the MAF. The extract used in this analysis may differ from the full Master Address File.

¹⁰ A small number of Individual Tax Identification Numbers (ITINs) were in the reference file when a PIK was assigned to 2010 Census persons. Additionally, ITINs were in the reference file when a PIK was assigned to some of the administrative data sources.

pairs as the 2010 Census, then the count ratio is 100 percent. Count ratios above 100 percent indicate a higher count in administrative records, while a ratio below 100 percent indicates a lower count in administrative records. Count ratios closer to 100 percent indicate better administrative data whereas very low and very high count ratios indicate lower quality administrative data.

The match ratio is calculated by dividing the count of 2010 Census records that match to administrative records by the 2010 Census count and multiplying the result by 100. The match ratio represents the percentage of 2010 Census addresses, persons, person-address pairs, and demographic characteristics that match to administrative records by MAFID, PIK, and PIK-MAFID, respectively.

3.4 Best Address for Person-Address Pairs

Administrative data sometimes have conflicting information regarding person-address pairs. For instance, one data source could have a person living at an address in Maryland, while another data source may have the same person living in Texas. To compare administrative records to the 2010 Census, a best address was chosen for persons with multiple addresses in administrative records.

A logistic model was utilized to select the best address for a person-address pair. For each administrative records source, the model estimated whether a particular administrative record address is the same as the 2010 Census address for each person found in both the 2010 Census and administrative records. The independent variables were 2010 Census demographic characteristics and proximity of an administrative record to April 1, 2010. Predicted values were obtained from each regression. For each person, the address associated with the highest predicted probability of having the same administrative records and 2010 Census address was selected. When demographic characteristics for a person were unavailable, the address was selected from the source with the highest overall address match rate with the 2010 Census.

There are persons at multiple addresses in the 2010 Census as well (when the same PIK appears at multiple MAFIDs), but for the person-address section these possible duplicates were kept in the 2010 Census universe.¹¹

4. Limitations

The 2010 Census Match Study included validated addresses and persons. Records lacking complete or quality data to match to the MAF or the person reference file were omitted from most analyses. The person reference file was based on the SSA Numident file which primarily includes persons with a SSN.

¹¹ The 2010 Census duplicates were retained in the count and match analyses pending further analysis on whether the pairs were true duplicates or error resulting from the probabilistic matching in the PIK assignment process.

One of the goals of the 2010 Census Match Study was to evaluate all items on the 2010 Census, including tenure and relationship to the householder. The administrative records data used in this study did not have tenure or relationship information on the files. Future research should evaluate how previous census records compare to the 2010 Census tenure and relationship data.

The majority of the federal and commercial data do not include group quarters, while the 2010 Census has housing units and group quarters. This report does not distinguish between those who live in group quarters and those who live in housing units in the 2010 Census.

5. Results

5.1 Address Count and Match

Nation

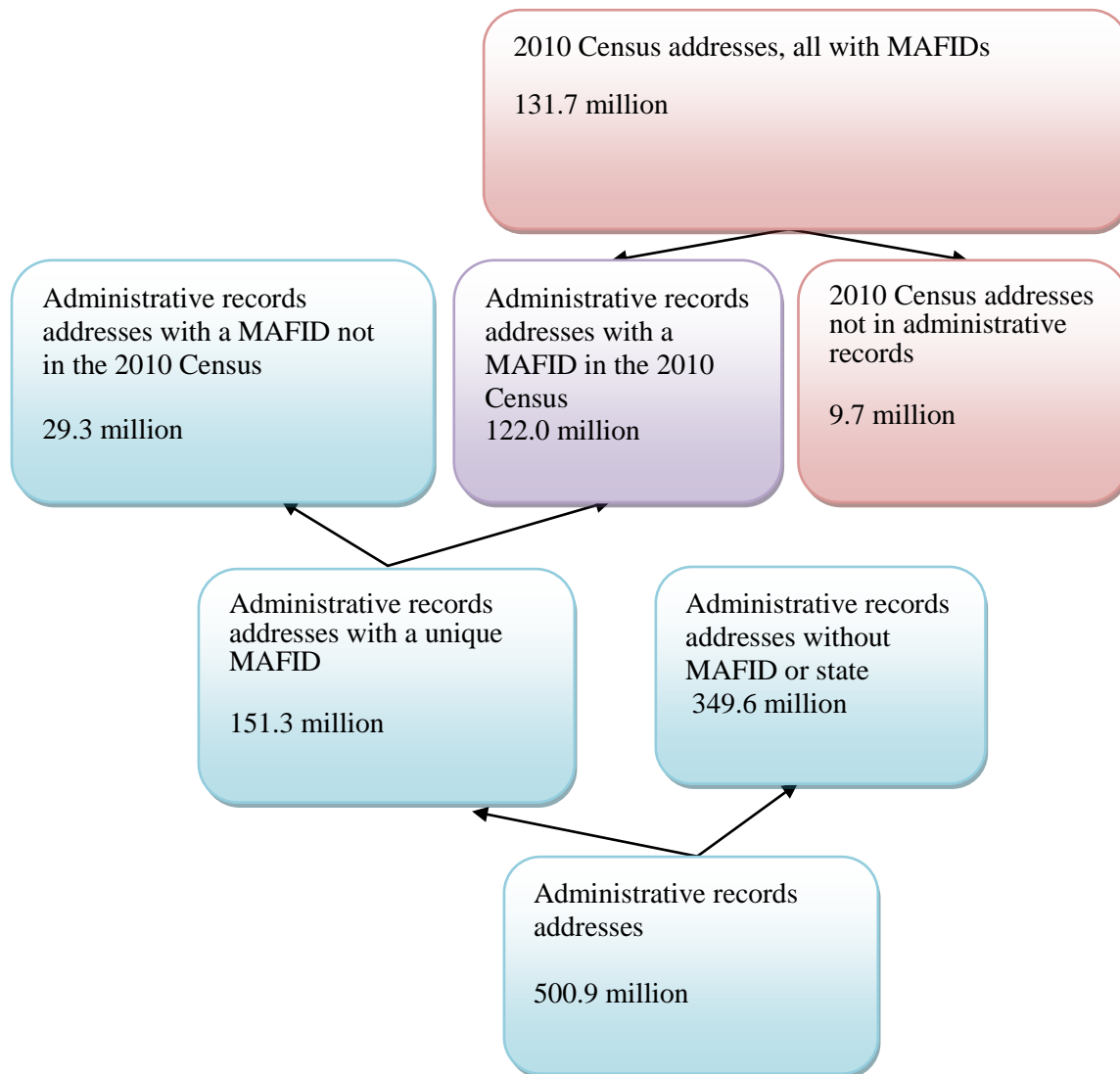
Figure 1 displays the number of addresses in the 2010 Census and administrative records. As discussed in the methodology section, MAFIDs are unique identifiers for addresses. For this report, MAFIDs facilitated address record linkage between the 2010 Census and administrative records.

There were 131.7 million occupied or vacant addresses in the 2010 Census, all of which had MAFIDs. There were 500.9 million addresses in the administrative records files. Of these, there were 151.3 million addresses that had a unique MAFID and 349.6 million addresses that did not have a MAFID. Future research will investigate unduplicating and assigning MAFIDs to administrative records addresses that do not have a MAFID.

Of the 131.7 million 2010 Census addresses, 122.0 million (92.6 percent) matched to administrative records addresses with MAFIDs. There were 29.3 million administrative records addresses with MAFIDs that were not in the 2010 Census and 9.7 million addresses that were in the 2010 Census, but not in administrative records.

There are several factors that impact the 2010 Census and administrative records address counts and matches. The 2010 Census addresses were physical locations, whereas administrative record data represented mailing addresses. For instance, there were Post Office (P.O.) Box addresses in administrative data, while the 2010 Census did not include P.O. Box addresses. Also, the 2010 Census included physical descriptions of addresses such as “yellow house near fork in the road,” which cannot be matched to administrative data. In addition, some of the commercial data utilized in this report included current and historical addresses, thus potentially containing old addresses that did not exist in April 2010.

Figure 1. Count and Match of 2010 Census and Administrative Records Addresses



Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The commercial data also included non-residential addresses. Preliminary research suggests that of the 29.3 million addresses in administrative records that were not in the 2010 Census, approximately 10.1 million may have been non-residential addresses (Schellhamer 2012). The administrative records data also could have contained addresses that were unknown to the Census Bureau such as new construction. These factors that contribute to the count and match differentials between the 2010 Census and administrative records will be examined further, contributing to research for the 2020 Census.

These results compare addresses with MAFIDs in administrative records to MAFIDs deemed “good census addresses” through 2010 Census operations. Additional research is required to

determine whether the universe of administrative records addresses could have been further refined.

Region

Table 1 shows the count and match results comparing the 2010 Census addresses to administrative records addresses by region.¹²

Table 1. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by Region

Region	2010 Census Address Count	Administrative Records Address Count	2010 Census and Administrative Records Address Match	2010 Census and Administrative Records Address Count Ratio	2010 Census and Administrative Records Address Match Ratio
Total	131,704,730	151,277,043	121,967,283	114.9	92.6
Northeast	23,647,636	26,090,251	21,410,938	110.3	90.5
Midwest	29,483,646	33,826,863	27,851,765	114.7	94.5
South	49,980,829	59,002,109	46,166,891	118.0	92.4
West	28,592,619	32,357,820	26,537,689	113.2	92.8

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

As discussed above, there were 131.7 million addresses in the 2010 Census and 151.3 million addresses in administrative records that received a MAFID, resulting in 19.6 million more addresses in administrative records relative to the 2010 Census. The address count ratio for the total population was 114.9 percent, mirroring the counts, which indicated a larger number of administrative records addresses relative to the 2010 Census.

All regions had count ratios above 110.0 percent.¹³ The South had the highest count ratio at 118.0 percent. In the South, there were 50.0 million addresses in the 2010 Census and 59.0 million in administrative records. The Midwest had the second highest count ratio (114.7 percent), where the 2010 Census count was 29.5 million and the administrative records count was 33.8 million. The count ratio for the West was 113.2 percent, and the Northeast had the lowest count ratio at 110.3 percent.

¹² Geographic variables in 2010 Census data were used to tabulate region, state, and county tables and figures throughout this report.

¹³ The Northeast census region includes Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Midwest census region includes Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. The South census region includes Alabama, Arkansas, Delaware, the District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia. The West census region includes Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.

As discussed above, the address match ratio for the total population was 92.6 percent, and all regions had a match ratio above 90.0 percent. The Midwest had the highest match ratio at 94.5 percent. Of the 29.5 million addresses in the Midwest in the 2010 Census, administrative records matched to 27.9 million. The West had the second highest match ratio (92.8 percent), followed by the South (92.4 percent). The Northeast had the lowest match ratio (90.5 percent). As demonstrated by the regional pattern of count and match ratios for addresses, these ratios do not necessarily correspond to each other. The Northeast had the lowest count and match ratios of all regions, while the South had the highest count ratio, but the second lowest match ratio.

State

Table 2 shows count ratios, match ratios, and the distribution of Type of Enumeration Area (TEA) for the ten states that have the lowest and highest count and match ratios (see Appendix 1 for 2010 Census and administrative records address count and match numbers and ratios for all states).

The state-level address count ratio ranged from 92.7 percent to 124.0 percent. Consistent with the finding that the South had the highest address count ratio relative to the other regions, many of the states with the highest count ratios are located in the South. Mississippi had the highest count ratio (124.0 percent), followed by Delaware (122.7 percent), Georgia (121.8 percent), Alabama (121.1 percent), and Louisiana (120.3 percent). All of these states are located in the South and of the ten states that had the highest count ratios, Iowa was the only one not in the South.

The state with the lowest count ratio was Alaska at 92.7 percent. This was the only state where the count ratio was below 100.0 percent. After Alaska, West Virginia (103.9 percent), Vermont (106.8 percent), Maine (106.8 percent), and New York (107.0 percent) had the next lowest state count ratios. Of the ten states that had the lowest count ratios, five were in the Northeast, four were in the West, and one was in the South. This is consistent with the regional patterns observed for count ratios, where the West and Northeast had lower count ratios relative to the South and Midwest.

Table 2. 2010 Census and Administrative Records Address Count Ratio, Match Ratio, and Type of Enumeration Area for the Ten States with the Lowest and Highest Ratios

State	Ratio	Type of Enumeration Area						
		Mailout / Mailback	Military	Remote Alaska	Remote Update Enumerate	Update Enumerate	Update / Leave	Urban Update / Leave
Lowest Count Ratios								
Alaska	92.7	63.2	2.3	9.3	1.3	0.0	23.9	0.0
West Virginia	103.9	54.9	0.0	0.0	0.0	0.0	45.1	0.0
Vermont	106.8	66.5	0.0	0.0	0.0	6.4	27.1	0.0
Maine	106.8	66.4	0.1	0.0	0.4	2.2	30.9	0.0
New York	107.0	93.9	0.1	0.0	0.0	1.9	4.1	0.0
Wyoming	107.8	51.4	0.4	0.0	0.0	4.0	44.2	0.0
New Mexico	108.0	66.6	0.4	0.0	0.0	6.4	26.5	0.0
Montana	108.3	40.6	0.2	0.0	0.0	6.3	52.9	0.0
New Hampshire	110.0	77.3	0.0	0.0	0.0	4.6	18.1	0.0
Rhode Island	110.0	98.3	0.3	0.0	0.0	0.4	1.0	0.0
Highest Count Ratios								
Mississippi	124.0	79.4	0.2	0.0	0.0	0.0	1.5	18.9
Delaware	122.7	89.4	0.4	0.0	0.0	0.0	10.3	0.0
Georgia	121.8	93.8	0.2	0.0	0.0	0.0	0.4	5.6
Alabama	121.1	88.4	0.1	0.0	0.0	1.2	0.3	9.9
Louisiana	120.3	68.9	0.3	0.0	0.0	0.0	1.3	29.5
Arkansas	118.8	66.9	0.1	0.0	0.0	0.0	33.0	0.0
Tennessee	118.6	99.6	0.1	0.0	0.0	0.3	0.0	0.1
Iowa	118.6	77.2	0.0	0.0	0.0	0.3	22.5	0.0
Texas	118.3	88.7	0.2	0.0	0.0	1.3	1.5	8.3
Florida	118.2	96.6	0.1	0.0	0.0	1.7	0.1	1.6
Lowest Match Ratios								
Alaska	70.5	63.2	2.3	9.3	1.3	0.0	23.9	0.0
West Virginia	72.8	54.9	0.0	0.0	0.0	0.0	45.1	0.0
Vermont	79.9	66.5	0.0	0.0	0.0	6.4	27.1	0.0
Maine	80.5	66.4	0.1	0.0	0.4	2.2	30.9	0.0
Montana	81.1	40.6	0.2	0.0	0.0	6.3	52.9	0.0
New Mexico	81.9	66.6	0.4	0.0	0.0	6.4	26.5	0.0
Wyoming	84.5	51.4	0.4	0.0	0.0	4.0	44.2	0.0
Hawaii	85.9	69.9	3.6	0.0	0.0	0.0	26.5	0.0
New Hampshire	87.5	77.3	0.0	0.0	0.0	4.6	18.1	0.0
Idaho	87.8	82.6	0.2	0.0	0.0	3.1	14.1	0.0
Highest Match Ratios								
Iowa	96.5	77.2	0.0	0.0	0.0	0.3	22.5	0.0
Ohio	96.1	96.7	0.0	0.0	0.0	0.0	3.3	0.0
District of Columbia	96.1	99.7	0.3	0.0	0.0	0.0	0.0	0.0
Maryland	95.8	98.4	0.3	0.0	0.0	0.0	1.4	0.0
Indiana	95.7	95.0	0.0	0.0	0.0	0.0	5.0	0.0
Kansas	95.4	80.9	0.5	0.0	0.0	0.2	18.4	0.0
California	95.4	96.5	0.2	0.0	0.0	0.9	2.2	0.1
Connecticut	95.0	99.3	0.1	0.0	0.0	0.0	0.5	0.0
Florida	94.8	96.6	0.1	0.0	0.0	1.7	0.1	1.6
Nebraska	94.6	77.1	0.3	0.0	0.0	0.4	22.3	0.0

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The state address match ratio range was 70.5 percent to 96.5 percent. The five states with the highest match ratios were in the Midwest and South. Iowa had the highest address match ratio at 96.5 percent. Ohio had the second highest match (96.1 percent), followed by the District of Columbia (96.1 percent), Maryland (95.8 percent), and Indiana (95.7 percent).¹⁴ Of the ten states that had the highest match ratios, five were in the Midwest, three in the South, and one each in the Northeast and West. Alaska (70.5 percent) had the lowest percent of addresses that matched between the 2010 Census and administrative records with MAFIDs. The following four states had the next lowest match ratios: West Virginia (72.8 percent), Vermont (79.9 percent), Maine (80.5 percent), and Montana (81.1 percent). Of the ten states with the lowest match ratios, six were in the West, three in the Northeast, and one in the South.

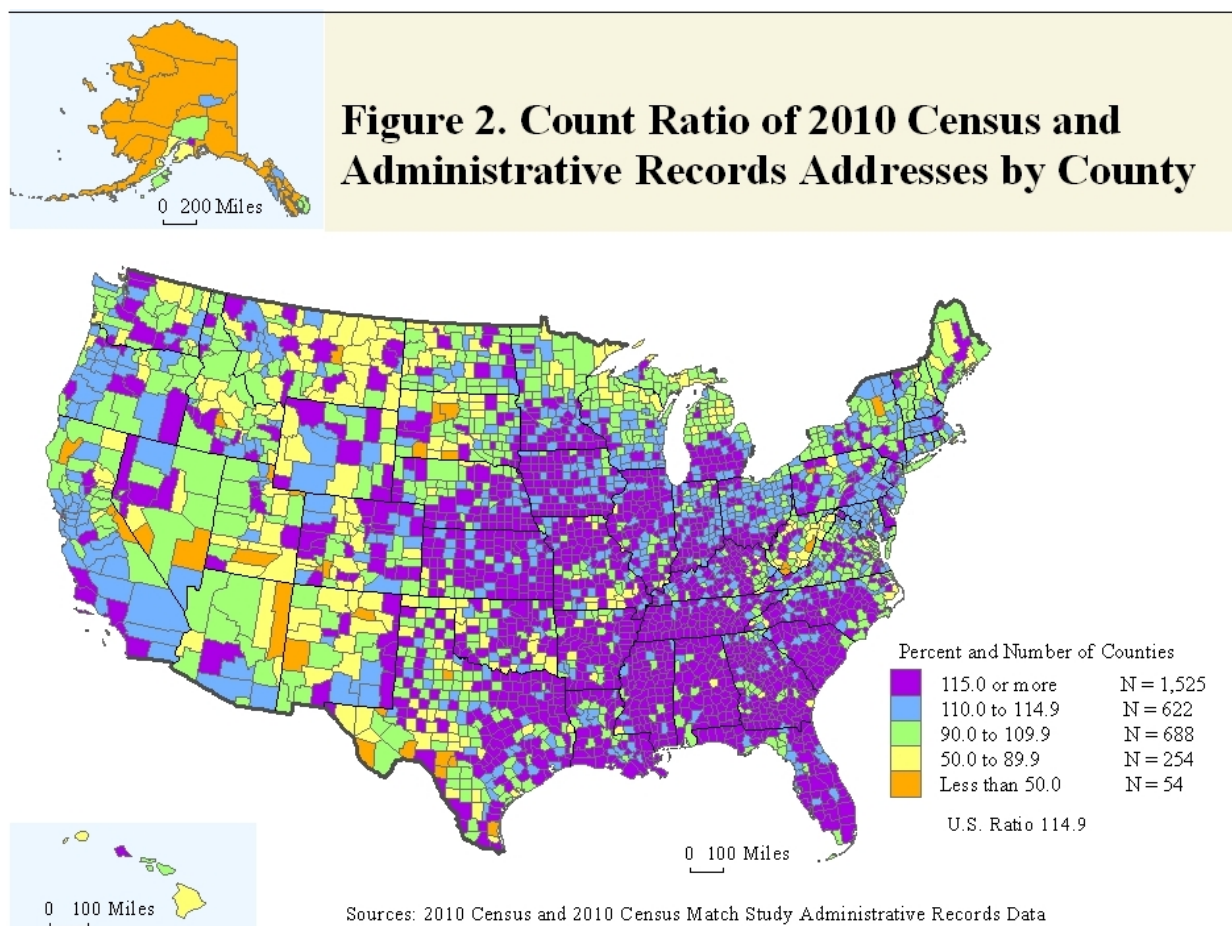
Future research will identify reasons behind geographic differences in count and match ratios. For instance, 2010 Census and administrative records address counts and matches may be in part affected by differences in city-style and rural route addresses, where city-style addresses are easier to match. TEA can be used as an indicator of city-style addresses as compared to incomplete or rural route addresses, as Mailout/Mailback TEAs tend to have more city-style addresses relative to other TEAs such as Update/Leave and Update Enumerate.¹⁵ Looking at the ten states in Table 2 with the lowest and highest count and match ratios by TEA, many of the states with the lowest count and match ratios had lower proportions of addresses designated as the Mailout/Mailback TEA relative to states with the highest count and match ratios.

County

Figure 2 shows address count ratios for the 2010 Census and administrative records by county. Green indicates counties with a count ratio that is closer to 100.0 percent, yellow and orange indicate low count ratios, and blue and purple represent high count ratios.

¹⁴ For this report, the District of Columbia is treated as a state equivalent.

¹⁵ The Census Bureau assigns a Type of Enumeration Area (TEA) value to collection blocks to assist with planning census operations for the decennial censuses. For instance, areas that have confirmed mail delivery by the U.S. Postal Service and good response rates to data collection efforts are generally assigned to a Mailout/Mailback TEA (Johanson et al. 2011). Mailout/Mailback is a data collection where forms are mailed to housing units and respondents are asked to complete their form and return by mail. Other TEAs include Update/Leave, Remote Update Enumerate, Remote Alaska, Update Enumerate, Military, and Urban Update/Leave. Update/Leave is a form of data collection where enumerators deliver questionnaires to housing units in their assignment areas and respondents are asked to complete their forms and return by mail. In Remote Update Enumerate, enumerators enumerate households; this is done in rural areas that may require special travel. Remote Alaska is a data collection method in isolated parts of Alaska where an enumerator enumerates the household. Update Enumerate is a data collection method for communities that have special needs, where an enumerator collects data from the household. Military represents areas that have military installations. Mailout/Mailback is conducted in these areas. Urban Update/Leave is a data collection method conducted in areas that have city-style addresses, but may not have good mail delivery. Enumerators leave questionnaires at housing units in their assignment areas and respondents are asked to complete and return the forms by mail. For more information on TEA delineation and definitions for the 2010 Census see Johanson et al. 2011.



Consistent with regional and state descriptive statistics, many counties in states in the South and Midwest had count ratios above 110.0 percent, indicating that administrative records had a higher number of addresses in these counties relative to the 2010 Census. In the Midwest, states or areas with a number of counties with low count ratios included North Dakota, South Dakota, northern Minnesota, northern Wisconsin, and northern Michigan. In the South, West Virginia and Texas had a number of counties with low count ratios. The West had many counties with low count ratios. For instance, many counties in Alaska, Montana, and New Mexico had low count ratios.

Table 3 shows count ratios (upper panel), match ratios (lower panel), and TEA for selected counties.

Table 3. 2010 Census and Administrative Records Address Count Ratio, Match Ratio, and Type of Enumeration Area for the Ten Counties with the Lowest and Highest Ratios

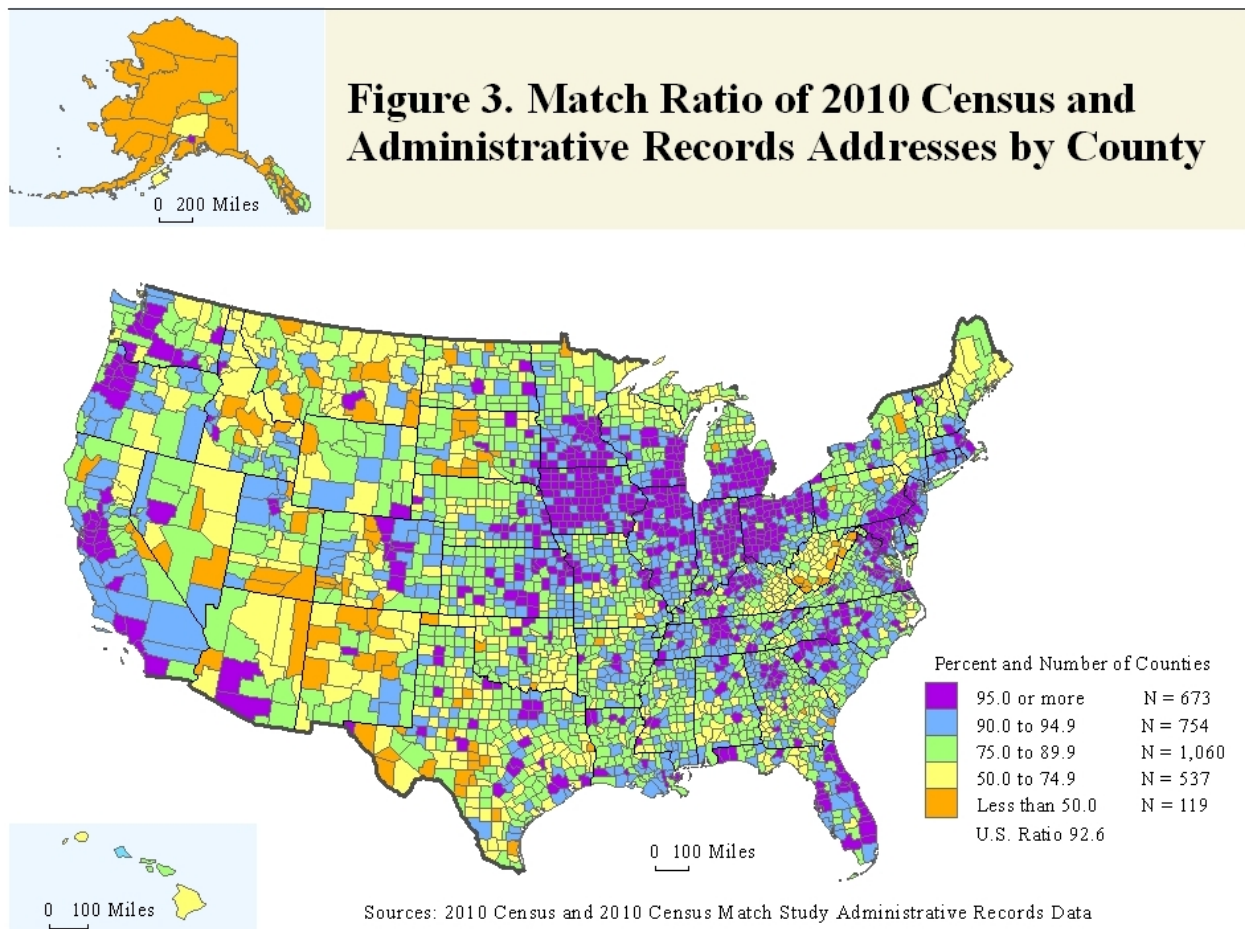
County		Type of Enumeration Area						
		Mailout / Mailback	Military	Remote Alaska	Remote Update Enumerate	Update Enumerate	Update / Leave	Urban Update / Leave
Ratio								
Lowest Count Ratios								
Denali, Alaska	12.1	0.0	0.2	22.8	0.0	0.0	77.0	0.0
Yukon-Koyukuk, Alaska	12.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Aleutians East, Alaska	14.7	0.0	0.0	50.2	0.0	0.0	49.8	0.0
Hoonah-Angoon, Alaska	15.5	0.0	0.0	0.0	21.2	0.0	78.8	0.0
Kalawao, Hawaii	15.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Yakutat, Alaska	16.0	0.0	0.0	0.0	14.9	0.0	85.1	0.0
Northwest Arctic, Alaska	18.4	0.0	0.0	57.1	0.0	0.0	42.9	0.0
Kenedy, Texas	19.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Dillingham, Alaska	20.2	0.0	0.0	56.9	0.0	0.0	43.1	0.0
Bethel, Alaska	21.2	0.0	0.0	60.1	0.0	0.0	39.9	0.0
Highest Count Ratios								
Chattahoochee, Georgia	208.2	0.0	55.6	0.0	0.0	0.0	0.0	44.4
Stephens, Georgia	207.2	66.9	0.0	0.0	0.0	0.0	0.0	33.1
Kiowa, Kansas	201.4	13.0	0.0	0.0	0.0	0.0	87.0	0.0
Hayes, Nebraska	185.3	0.0	0.0	0.0	0.0	0.0	100.0	0.0
St. Bernard, Louisiana	183.9	0.0	0.7	0.0	0.0	0.0	0.0	99.3
Warren, Georgia	180.6	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Clay, Tennessee	177.0	86.8	0.0	0.0	0.0	0.0	0.0	13.2
Houston, Tennessee	176.5	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Bleckley, Georgia	170.5	30.5	0.0	0.0	0.0	0.0	0.0	69.5
Pike, Georgia	169.7	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Lowest Match Ratios								
Denali, Alaska	8.0	0.0	0.2	22.8	0.0	0.0	77.0	0.0
Yukon-Koyukuk, Alaska	9.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Aleutians East, Alaska	12.2	0.0	0.0	50.2	0.0	0.0	49.8	0.0
Kalawao, Hawaii	12.4	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Aleutians West, Alaska	12.8	0.0	0.0	42.7	0.0	0.0	57.3	0.0
Yakutat, Alaska	12.9	0.0	0.0	0.0	14.9	0.0	85.1	0.0
Shannon, South Dakota	13.7	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Dillingham, Alaska	13.9	0.0	0.0	56.9	0.0	0.0	43.1	0.0
Haines, Alaska	14.5	0.0	0.0	0.0	0.1	0.0	99.9	0.0
Northwest Arctic, Alaska	14.9	0.0	0.0	57.1	0.0	0.0	42.9	0.0
Highest Match Ratios								
Manassas Park, Virginia	99.5	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Alexandria, Virginia	99.1	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Manassas, Virginia	98.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0
St. Louis, Missouri	98.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Anoka, Minnesota	98.8	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Radford, Virginia	98.7	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Hampton, Virginia	98.7	99.0	1.0	0.0	0.0	0.0	0.0	0.0
Los Alamos, New Mexico	98.6	94.0	0.0	0.0	0.0	0.0	6.0	0.0
Fairfax, Virginia	98.6	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Minnehaha, South Dakota	98.6	100.0	0.0	0.0	0.0	0.0	0.0	0.0

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The count ratio range across counties was 12.1 percent to 208.2 percent. Of the ten counties with the lowest count ratios, eight were in Alaska—Denali (12.1 percent), Yukon-Koyukuk (12.1 percent), Aleutians East (14.7 percent), Hoonah-Angoon (15.5 percent), Yakutat (16.0 percent), Northwest Arctic (18.4 percent), Dillingham (20.2 percent), and Bethel (21.2 percent). Kalawao, Hawaii (15.9 percent) and Kenedy, Texas (19.7 percent) were also among the ten counties with the lowest count ratios.

The ten counties that had the highest count ratios were mainly in the South, many of them in Georgia—Chattahoochee (208.2 percent), Stephens (207.2 percent), Warren (180.6 percent), Bleckley (170.5 percent), and Pike (169.7 percent). Two counties were in Tennessee—Clay (177.0 percent) and Houston (176.5 percent). St. Bernard, Louisiana (183.9 percent) was also among the ten counties with the highest count ratios. Two counties in the Midwest, Kiowa, Kansas (201.4 percent) and Hayes, Nebraska (185.3 percent) were also among the top ten.

Figure 3 shows address match ratios by county. Purple and blue represent counties with high match ratios, while yellow and orange represent low match ratios.



A number of states in the Midwest had counties with high match ratios. The majority of counties in Iowa, Illinois, Indiana, and Ohio had match ratios that were 90.0 percent or above. Many counties in southern Minnesota, Wisconsin, and Michigan also had match ratios that were 90.0 percent or above. In the Northeast, New Jersey, southeast Pennsylvania, Connecticut, Rhode Island and Massachusetts had a number of counties with high match ratios.

Across the United States, counties near metropolitan areas tended to have high match ratios. For instance, in the West, counties near Los Angeles, California; San Francisco, California; Portland, Oregon; Seattle, Washington; and Denver, Colorado had high match ratios. In the South, counties near Houston, Texas; Austin, Texas; Little Rock, Arkansas; Birmingham, Alabama; Montgomery, Alabama; and Atlanta, Georgia had high match ratios. Many counties in western states, such as Alaska, Idaho, Montana, and New Mexico had a number of counties with low match ratios. West Virginia in the South was another state that had many counties with low match ratios.

Address match ratios for counties ranged from 8.0 percent to 99.5 percent (Table 3, bottom panel). Similar to the address count ratios for counties, eight of the ten counties with the lowest match ratios were in Alaska—Denali (8.0 percent), Yukon-Koyukuk (9.1 percent), Aleutians East (12.2 percent), Aleutians West (12.8 percent), Yakutat (12.9 percent), Dillingham (13.9 percent), Haines (14.5 percent), and Northwest Arctic (14.9 percent). Shannon, South Dakota (13.7 percent) and Kalawao, Hawaii (12.4 percent) were also among the ten counties with the lowest match ratios.

The ten counties with the highest match ratios were located in the South, Midwest and West. Six counties from Virginia were within the ten counties with the highest match ratios, Manassas Park (99.5 percent), Alexandria (99.1 percent), Manassas (98.9 percent), Radford (98.7 percent), Hampton (98.7 percent), and Fairfax (98.6 percent). Three counties in the Midwest were among the ten counties with the highest match ratios, St. Louis, Missouri (98.9 percent); Anoka, Minnesota (98.8 percent); and Minnehaha, South Dakota (98.6 percent). Los Alamos, New Mexico also had a high address match ratio (98.6 percent).

Similar to the state patterns, TEAs explain some of the count and match trends by county. Counties with the lowest count and match ratios did not have any addresses in the Mailout/Mailback TEA. Eight of the ten counties with the highest match ratios were entirely in the Mailout/Mailback TEA. At least 94.0 percent of the 2010 Census addresses in the other two counties with the highest match ratios were in Mailout/Mailback.

Federal and Commercial Data

Table 4 shows count and match ratios for the 2010 Census and federal and commercial data.

Table 4. 2010 Census and Federal and Commercial Administrative Records Address Count and Match Numbers and Ratios

Data Type	Administrative Records Address Count	2010 Census and Administrative Records Address Match	2010 Census and Administrative Records Count Ratio	2010 Census and Administrative Records Address Match Ratio
Commercial	145,635,096	119,035,878	110.6	90.4
Federal	122,680,039	110,914,836	93.1	84.2
In both Commercial and Federal	117,038,092	107,983,431	88.9	82.0

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

There were more unduplicated addresses with MAFIDs in commercial data compared to both the 2010 Census and federal data. There were 145.6 million addresses in commercial data and 122.7 million addresses in federal data. There were 117.0 million addresses that were in both commercial and federal data. Thus, 28.6 million addresses were unique to commercial data, and 5.6 million addresses were unique to federal data.

The 2010 Census-commercial and 2010 Census-federal count ratios were 110.6 percent and 93.1 percent respectively. Commercial data not only had a higher 2010 Census address count ratio relative to federal data, they also had higher 2010 Census address match ratios. Of the 131.7 million addresses in the 2010 Census, commercial addresses matched to 119.0 million or 90.4 percent. Federal data matched to 110.9 million or 84.2 percent of the 2010 Census addresses.

Type of Enumeration Area

Table 5 shows TEA address count and match ratios for the 2010 Census and administrative records.^{16, 17}

TEAs that were designated for Mailout/Mailback data collection methods, where forms were mailed to housing units and respondents were asked to complete and mail back their questionnaire, had the highest count and match ratios—Mailout/Mailback and Military TEAs. The Mailout/Mailback TEA had the second highest count ratio and highest match ratio at 114.1 percent and 94.6 percent respectively. The Military TEA had the highest count ratio and second highest match ratio at 200.5 percent and 92.8 percent respectively. These TEAs were designated for Mailout/Mailback data collection in part because they had confirmed mail delivery by the postal service and had fewer enumeration challenges (Johanson et al. 2011). These addresses were also mostly city-style addresses, which generally pose less of a matching issue relative to rural route addresses (Johanson et al. 2011).

¹⁶ Note that counts for TEA differ from “2010 Census Operational Assessment for Type of Enumeration Area Delineation” (Johanson et al. 2011) as different data sets were used.

¹⁷ Not all administrative records addresses were assigned a TEA as these may include new construction that did not exist prior to address canvassing as well as non-residential addresses which are not assigned a TEA.

Table 5. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by Type of Enumeration Area

Type of Enumeration Area	2010 Census Address Count	Administrative Records Address Count	2010 Census and Administrative Records Address Match	2010 Census and Administrative Records Address Count Ratio	2010 Census and Administrative Records Address Match Ratio
Total	131,704,730	151,277,043	121,967,283	114.9	92.6
Mailout/Mailback	119,713,726	136,634,851	113,204,798	114.1	94.6
Military	213,420	427,947	198,082	200.5	92.8
Remote Alaska	28,549	5,710	4,798	20.0	16.8
Remote Update Enumerate	6,896	2,481	1,898	36.0	27.5
Update Enumerate	1,366,883	1,149,847	875,505	84.1	64.1
Update/Leave	7,978,221	7,571,640	5,863,855	94.9	73.5
Urban Update/Leave	2,397,035	2,497,466	1,818,347	104.2	75.9
No TEA	0	2,987,101	0	-	-

Note: A “-” in tables in this report indicates a ratio where the denominator was 0.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The Urban Update/Leave TEA had the next highest count and match ratios, but they were considerably lower than the Mailout/Mailback and Military count and match ratios. The count ratio was 104.2 percent, and the match ratio was 75.9 percent. These count and match ratios were likely lower for Urban Update/Leave relative to Mailout/Mailback and Military because this form of data collection was designated for areas where the Census Bureau believed that there were issues with accurate mail delivery (Johanson et al. 2011). For instance, this TEA included multi-unit buildings where mail was delivered at a drop point instead of individual units or communities that had city-style addresses, but where many residents had mail delivered to a P.O. Box, likely impacting the match ratios between the 2010 Census and administrative records. (Johanson et al. 2011).

The Update/Leave count ratio (94.9 percent) was lower than Urban Update/Leave, but this TEA had a similar match ratio (73.5 percent) to Urban Update/Leave. The Update Enumerate count and match ratios were 84.1 percent and 64.1 percent respectively. Update/Leave and Update Enumerate count and match ratios were likely lower than Mailout/Mailback and Military because Update/Leave was conducted in areas that typically do not have city-style addresses and in Update Enumerate many housing units may not have had a house number or street name, making these addresses difficult to match (Johanson et al. 2011).

Remote Alaska and Remote Update Enumerate had the lowest count and match ratios. The count and match ratios for Remote Update Enumerate were 36.0 percent and 27.5 percent respectively. Remote Alaska had the lowest count and match ratios at 20.0 percent and 16.8 percent respectively. Remote Alaska and Remote Update Enumerate areas were designated as such because mail was considered undeliverable, thus accounting for the low count and match ratios.

Housing Unit Type

Table 6 shows 2010 Census and administrative records address count and match ratios by housing unit type.¹⁸ Multi-unit buildings with five to nine units had the highest count ratio (145.5 percent), followed by buildings with ten to nineteen units (136.6 percent). Multi-unit buildings with 20 or more units (118.7 percent), multi-unit buildings with two to four units (115.9 percent), and single-family homes (114.2 percent) all had lower count ratios, but they were still above 100 percent. The count ratio for trailer-mobile homes was considerably lower at 88.7 percent. The category “other,” which includes boats, recreational vehicles, and vans had the lowest count ratio at 49.4 percent.

Table 6. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by Housing Unit Type

Housing Unit Type	2010 Census Address Count	Administrative Records Address Count	2010 Census and Administrative Records Address Match	2010 Census and Administrative Records Address Count Ratio	2010 Census and Administrative Records Address Match Ratio
Total	131,704,730	151,277,043	121,967,283	114.9	92.6
Multi-unit building – 2 to 4 units	7,412,416	8,590,969	5,902,065	115.9	79.6
Multi-unit building – 5 to 9 units	3,807,849	5,540,284	3,529,097	145.5	92.7
Multi-unit building – 10 to 19 units	4,069,731	5,559,212	3,814,398	136.6	93.7
Multi-unit building – 20 or more units	14,184,728	16,838,161	13,137,945	118.7	92.6
Other - boat, recreation vehicle, van, etc.	125,493	61,966	34,409	49.4	27.4
Single-family Home	94,744,173	108,158,255	89,506,322	114.2	94.5
Trailer-Mobile home	7,360,340	6,528,196	6,043,047	88.7	82.1

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Single-family homes had the highest housing type match ratio at 94.5 percent. Most of the addresses in the United States were single-family homes in the 2010 Census at 94.7 million. Of these, administrative records matched to 89.5 million. After single-family homes, the next highest matches were in multi-unit buildings with ten to nineteen units (93.7 percent), five to nine units (92.7 percent), and 20 or more units (92.6 percent), followed by trailer-mobile homes (82.1 percent).

Multi-unit buildings with two to four units (79.6 percent) had a considerably lower percentage match relative to other multi-unit building categories. This lower match may be in part due to smaller multi-unit structures having potentially more problematic addresses in some parts of the country. In some geographic areas, units are added to single units or small multi-units, and these added units may lack unit designations or mail may be delivered to one box (Virgile 2012).

¹⁸ 2010 Census and administrative records address housing unit type was assigned based on unit type designation in the MAF, the structure point permanent ID, and the number of units assigned to the MAFID in the MAF.

These types of situations would make addresses more difficult to match between administrative records and 2010 Census data.

Race and Hispanic Origin of Householder and Census Operations

Thus far, the figures and tables that have been discussed focus on the 131.7 million addresses in the 2010 Census and 151.3 million addresses in administrative records, regardless of whether they were occupied or vacant. The universe for this sub-section is occupied housing units.

Table 7 shows 2010 Census match ratios by 2010 Census race and Hispanic origin of the householder and mode of data collection. Count ratios are not included because administrative records address data did not include demographic characteristics on the householder. The occupied housing unit universe in Table 7 is 116.7 million housing units. Of the occupied housing units in the 2010 Census, administrative records matched to 110.5 million.

Note that the characteristic, mode, count imputation, and proxy data in Table 7 is from the 2010 Census, thus the Hispanic origin and race of householder analysis is not based on matched Hispanic origin and race responses in the 2010 Census and administrative records. Matched demographic response data will be evaluated in section 5.4 of this report.

The proportion of 2010 Census addresses that administrative records matched was similar for both Hispanic and non-Hispanic householders. Of the 13.5 million addresses that had a Hispanic householder, administrative records matched to 12.7 million or 94.2 percent. Of the 103.3 million addresses that had a non-Hispanic householder in the 2010 Census, administrative records matched to 97.8 million or 94.7 percent.

For race, 95.8 percent of 2010 Census addresses with a householder that reported Asian alone matched to administrative records. Of the 4.6 million addresses that had a householder that reported Asian alone, administrative records matched to 4.4 million. Addresses that had householders who reported Black alone had the next highest percentage matches at 94.9 percent, followed by White alone (94.8 percent), Two or More Races (94.3 percent), SOR alone (93.6 percent), and Native Hawaiian or Other Pacific Islander (NHPI) alone (93.5 percent). Addresses that had AIAN alone householders had a much lower match relative to the other race groups, 82.3 percent.

Table 7. 2010 Census and Administrative Records Address Match by Race and Hispanic Origin of Householder, Mode, Imputation, and Proxy

Demographic Characteristics of Householder, Mode, Count Imputation, and Proxy	2010 Census Address Count	2010 Census and Administrative Records Address Match	2010 Census and Administrative Records Address Match Ratio
Total Occupied Housing Units	116,716,292	110,504,340	94.7
Hispanic or Latino Origin			
Hispanic	13,461,366	12,681,754	94.2
Not Hispanic	103,254,926	97,822,586	94.7
Race			
White Alone	89,754,352	85,078,408	94.8
Black Alone	14,129,983	13,403,061	94.9
American Indian or Alaska Native Alone	939,707	773,742	82.3
Asian Alone	4,632,164	4,438,090	95.8
Native Hawaiian or Other Pacific Islander Alone	143,932	134,599	93.5
Some Other Race Alone	4,916,427	4,602,454	93.6
Two or More Races	2,199,727	2,073,986	94.3
Mode			
Nonresponse Followup	23,584,428	21,039,269	89.2
Mailout/Mailback	82,780,761	80,345,450	97.1
Other ¹	10,351,103	9,119,621	88.1
Count Imputation			
Not Imputed	116,282,183	110,166,897	94.7
Imputed	434,109	337,443	77.7
Proxy			
Not by Proxy	109,800,016	104,480,943	95.2
By Proxy	6,916,276	6,023,397	87.1

¹ The Mode category “Other” is a residual category that includes responses that were not obtained through either Nonresponse Followup or Mailout/Mailback.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The areas and associated TEAs where the AIAN population lives may in part be why the address match was lower in households with a householder who reported AIAN alone. For instance, only 70.2 percent of addresses where the householder was AIAN alone were in the Mailout/Mailback TEA compared to 90 percent and above for all other race groups. The match was conducted by MAFID; future research will explore whether the census addresses with high proportions of the AIAN population contained matchable addresses or physical locations.

Administrative records matched to 80.3 million or 97.1 percent of the 82.8 million 2010 Census addresses enumerated by Mailout/Mailback.¹⁹ A lower percentage of administrative records addresses matched to 2010 Census addresses in NRFU and other modes. Of the 23.6 million addresses that responded in NRFU, administrative records matched to 21.0 million or 89.2 percent. Of the approximately 434,000 addresses for which a population count was imputed, administrative records matched to approximately 337,000 or 77.7 percent. There were 6.9 million addresses that had a form of proxy response meaning that the 2010 Census response may have come from neighbors, building managers, or new households reporting on previous households. Of these, administrative records matched to 6.0 million or 87.1 percent. Since the quality of address data should not vary significantly between NRFU and Mailout/Mailback universes or between proxy and non-proxy cases, future research should further evaluate the address match ratio differences between Mailout/Mailback and NRFU and also proxy and non-proxy responses.

5.2 Person Count and Match

Nation

In this section, match ratios must be interpreted slightly differently compared to the previous section on addresses. In the address count and match section, all 2010 Census addresses had MAFIDs, therefore all of the 131.7 million addresses had the potential to be matched to administrative records with MAFIDs. This is not the case for persons, as not all persons in the 2010 Census received a unique person identifier, or PIK. This reduces the number of persons in the 2010 Census that have the potential to match to administrative records, contributing to lower match ratios for persons relative to addresses.

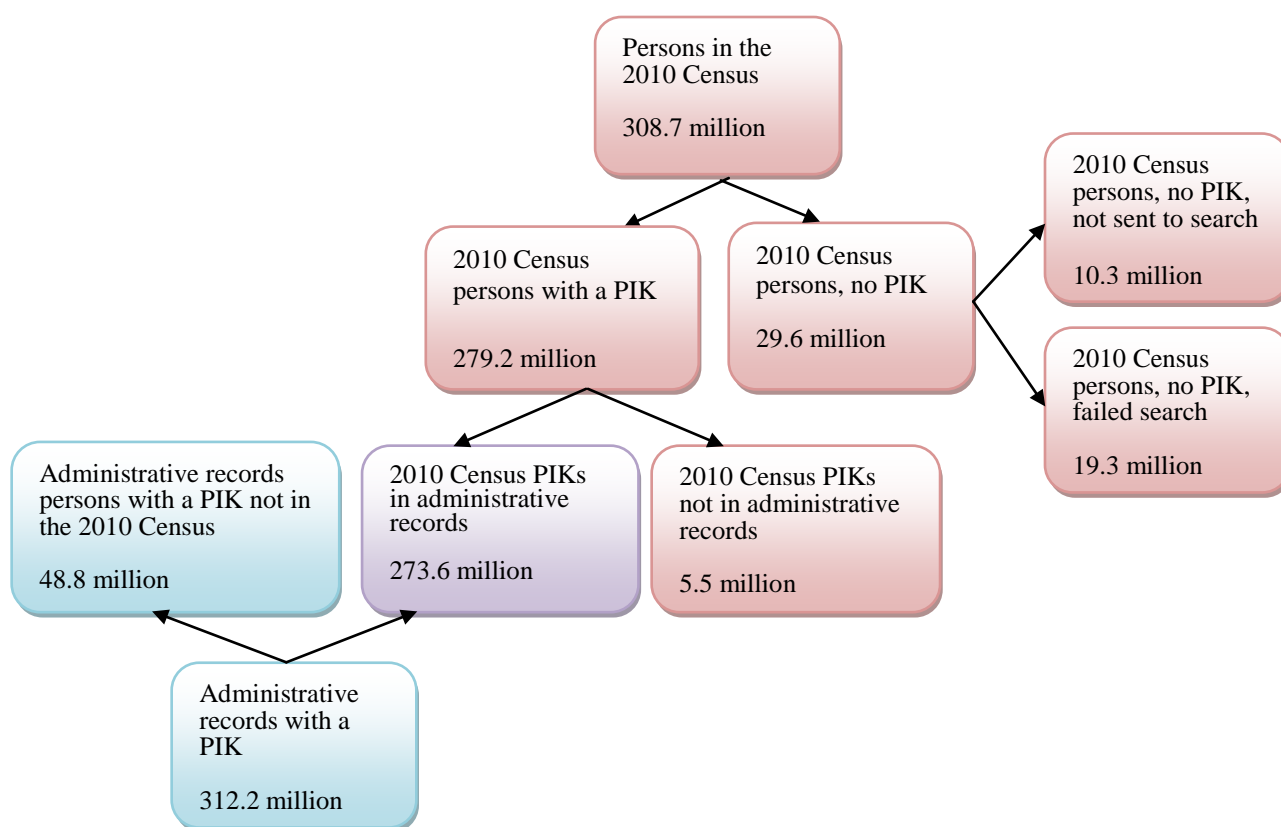
Figure 4 shows the number and match of 2010 Census and administrative records persons.²⁰ There were 308.7 million persons enumerated in the 2010 Census, 279.2 million of which had a

¹⁹ Mailout/Mailback for mode is different from Mailout/Mailback for TEA in this report, as the latter refers to collection blocks that are designated for Mailout/Mailback data collection in an effort to determine how to efficiently enumerate people living in various parts of the country, and the former refers to the mode by which the household was actually enumerated. For example, a household may be designated in a Mailout/Mailback TEA but that household may respond via Nonresponse Followup.

²⁰ The 2010 Census included duplicate PIKs whereas the administrative records contained unique PIKs. This resulted in instances where a single administrative record matched to multiple census records. Therefore, the sum of the count for persons in administrative records with a PIK not in the 2010 Census (48.8 million) and the count for 2010 Census PIKs in administrative records (273.6 million) does not equal the number of persons in administrative records with a PIK (312.2 million).

PIK. There were 312.2 million persons in administrative records that were alive on Census Day and had a PIK. Administrative records matched to the vast majority of 2010 Census PIKs, 273.6 million or 98.0 percent. The percentage of all 2010 Census persons, those with a PIK and those without, that matched to administrative records is about 10 percentage points lower at 88.6 percent. For the remainder of this section, unless otherwise specified, match ratios are based on the match of all persons in the 2010 Census relative to administrative records PIKs.

Figure 4. Count and Match of 2010 Census and Administrative Records Persons



Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

There were 29.6 million 2010 Census persons that failed the validation process and therefore did not receive a PIK, meaning that record linkage between these persons and administrative records was not possible.²¹ Of these 29.6 million persons, 10.3 million could not be sent through the PVS process as they lacked name and date of birth, and 19.3 million went through the PVS process but failed the validation process.

²¹ Future research will focus on direct matching of persons across files without validating against a reference file.

Of the 2010 Census unPIKed persons, there were 9.0 million persons whose 2010 Census response came from a form of proxy response wherein neighbors or new households reported on previous households. While many neighbors and new households may be able to provide name and date of birth information that would allow the validation of a record, other neighbors and new households were not able to provide this information adequately. There were also 1.2 million records that did not receive a PIK because the people in these households were imputed in the 2010 Census.

There were 48.8 million administrative records that were assigned a PIK but did not match to the 2010 Census. There is likely an overlap between the 29.6 million persons in the 2010 Census that did not receive a PIK and the 48.8 million persons in administrative records that received a PIK but did not match to the 2010 Census. Future research will study these two universes and their potential overlap.

There were 5.5 million 2010 Census persons with a PIK that were not in administrative records. About 4.0 million of these persons were children under the age of 17, and approximately 891,000 of these had an age of 0 in the 2010 Census. There are several reasons why this age group is less likely to be in administrative records compared to the 2010 Census. Tax data are one important source of information on children in administrative records. Therefore, how and when taxes are filed in combination with particular aspects of the tax data that the Census Bureau received from the IRS impact the coverage of children in administrative records. Children born on or after January 1, 2010 would not be claimed on 2009 taxes, therefore they may have been reported in the 2010 Census, but they would not likely be in the administrative records data used for this report. Additionally, tax forms such as 1040EZ do not collect data on dependents. There were also a number of dependents in administrative records that did not receive a PIK because there was not enough information to validate the records. Also, the IRS 1040 data used in the 2010 Census Match Study only had information on the first four dependents on a tax return, potentially limiting the number of children reported in larger households. Future research will include assessing other types of tax return data that include all dependents.

Region

Table 8 shows the 2010 Census person count, the number of PIKs in the 2010 Census, administrative records person count, the number of 2010 Census and administrative records that matched, and the 2010 Census and administrative records person count and match ratios by region.

The 2010 Census and administrative records person count ratio for the total U.S. population was 101.1 percent. All regions also had a person count ratio of about 101 percent. The Northeast and West had the same count ratio (101.4 percent). The Midwest had a slightly lower count ratio (101.1 percent), followed by the South (100.8 percent). These count ratios mirror the person

counts, where administrative records had a slightly higher count of persons relative to the 2010 Census for the total population and across all four regions.

Table 8. 2010 Census and Administrative Records Person Count and Match Numbers and Ratios by Region

	2010 Census Person Count	2010 Census Persons with a PIK	Administrative Records Person Count	2010 Census and Administrative Records Person Match	2010 Census and Administrative Records Person Count Ratio	2010 Census and Administrative Records Person Match Ratio
Total	308,745,538	279,179,329	312,214,325	273,643,411	101.1	88.6
Northeast	55,317,240	50,506,657	56,097,631	49,624,941	101.4	89.7
Midwest	66,927,001	62,498,752	67,672,118	61,340,240	101.1	91.7
South	114,555,744	102,720,450	115,504,373	100,766,768	100.8	88.0
West	71,945,553	63,453,470	72,940,203	61,911,462	101.4	86.1

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

As discussed above, 88.6 percent of all 2010 Census persons (PIKed and unPIKed) matched to administrative records. The Midwest had the highest percentage of 2010 Census persons that were PIKed (93.4 percent) and that matched to administrative records (91.7 percent). The Northeast had the second highest percentage of 2010 Census persons that were PIKed (91.3 percent) and that matched to administrative records (89.7 percent), followed by the South. The West had the lowest percentage of PIKed 2010 Census persons (88.2 percent) and 2010 Census records that matched to administrative records (86.1 percent).

State

Table 9 shows the 2010 Census person count, the number of PIKs in the 2010 Census, the administrative records person count, the number of 2010 Census and administrative records that matched, and the 2010 Census and administrative records count and match ratios by state.

The person count ratio ranged from 96.9 percent to 103.9 percent across states. Thirteen states had a count ratio below 100 percent, fifteen states had a count ratio of 100 percent, and twenty-three states had a count ratio greater than 100 percent. The states with the highest person count ratios were New Jersey (103.9 percent), Illinois (103.3 percent), Georgia (102.7 percent), California (102.4 percent), and Washington (102.4 percent). The states with the lowest person count ratios were Wyoming (96.9 percent), North Dakota (97.6 percent), Arizona (97.9 percent), Montana (98.2 percent), and New Mexico (98.3 percent). All of the regions were represented within the ten states that had the highest count ratios. Of the ten states with the lowest count ratios, half were in the West, two were in the South, two were in the Midwest, and one was in the Northeast.

Table 9. 2010 Census and Administrative Records Person Count and Match Numbers and Ratios by State

State	2010 Decennial Person Count	2010 Census Persons with a PIK	Administrative Records Person Count	2010 Census and Administrative Records Person Match	2010 Census and Administrative Records Person Count Ratio	2010 Census and Administrative Records Person Match Ratio
Total	308,745,538	279,179,329	312,214,325	273,643,411	101.1	88.6
Alabama	4,779,736	4,291,898	4,855,249	4,228,684	101.6	88.5
Alaska	710,231	640,013	716,305	635,613	100.9	89.5
Arizona	6,392,017	5,504,074	6,260,469	5,372,306	97.9	84.0
Arkansas	2,915,918	2,665,171	2,903,339	2,621,373	99.6	89.9
California	37,253,956	32,518,962	38,160,772	31,603,657	102.4	84.8
Colorado	5,029,196	4,482,335	5,039,949	4,391,915	100.2	87.3
Connecticut	3,574,097	3,307,240	3,608,268	3,253,223	101.0	91.0
Delaware	897,934	809,132	912,088	796,215	101.6	88.7
District of Columbia	601,723	522,688	606,137	511,746	100.7	85.0
Florida	18,801,310	16,800,443	19,008,662	16,493,170	101.1	87.7
Georgia	9,687,653	8,520,330	9,945,565	8,335,517	102.7	86.0
Hawaii	1,360,301	1,206,191	1,371,877	1,182,070	100.9	86.9
Idaho	1,567,582	1,428,711	1,546,532	1,397,038	98.7	89.1
Illinois	12,830,632	11,733,482	13,255,633	11,531,040	103.3	89.9
Indiana	6,483,802	6,054,511	6,572,141	5,958,989	101.4	91.9
Iowa	3,046,355	2,889,518	3,048,064	2,851,878	100.1	93.6
Kansas	2,853,118	2,670,501	2,873,274	2,624,387	100.7	92.0
Kentucky	4,339,367	3,994,765	4,397,339	3,934,626	101.3	90.7
Louisiana	4,533,372	4,065,851	4,583,043	3,996,293	101.1	88.2
Maine	1,328,361	1,256,619	1,340,538	1,239,680	100.9	93.3
Maryland	5,773,552	5,257,560	5,880,321	5,152,768	101.8	89.2
Massachusetts	6,547,629	6,087,938	6,651,229	5,990,853	101.6	91.5
Michigan	9,883,640	9,264,073	9,804,204	8,908,584	99.2	90.1
Minnesota	5,303,925	5,016,847	5,348,667	4,947,694	100.8	93.3
Mississippi	2,967,297	2,703,142	3,004,903	2,658,172	101.3	89.6
Missouri	5,988,927	5,578,535	5,987,199	5,493,569	100.0	91.7
Montana	989,415	902,296	971,295	890,441	98.2	90.0
Nebraska	1,826,341	1,705,041	1,832,976	1,681,487	100.4	92.1
Nevada	2,700,551	2,305,111	2,744,855	2,253,127	101.6	83.4
New Hampshire	1,316,470	1,244,718	1,335,435	1,228,380	101.4	93.3
New Jersey	8,791,894	7,976,238	9,138,823	7,836,027	103.9	89.1
New Mexico	2,059,179	1,783,742	2,023,747	1,749,475	98.3	85.0
New York	19,378,102	17,178,954	19,565,132	16,829,755	101.0	86.8
North Carolina	9,535,483	8,531,921	9,509,731	8,381,227	99.7	87.9
North Dakota	672,591	639,442	656,192	632,637	97.6	94.1
Ohio	11,536,504	10,811,996	11,740,953	10,654,439	101.8	92.4
Oklahoma	3,751,351	3,401,933	3,763,742	3,344,268	100.3	89.1
Oregon	3,831,074	3,485,866	3,868,850	3,422,049	101.0	89.3
Pennsylvania	12,702,379	11,893,542	12,779,595	11,704,799	100.6	92.1
Rhode Island	1,052,567	965,728	1,057,920	953,304	100.5	90.6
South Carolina	4,625,364	4,212,922	4,606,817	4,143,006	99.6	89.6
South Dakota	814,180	755,176	813,677	746,041	99.9	91.6
Tennessee	6,346,105	5,794,732	6,441,396	5,706,995	101.5	89.9
Texas	25,145,561	22,128,264	25,173,066	21,598,531	100.1	85.9
Utah	2,763,885	2,551,307	2,804,835	2,481,704	101.5	89.8
Vermont	625,741	595,680	620,691	588,920	99.2	94.1
Virginia	8,001,024	7,335,606	8,085,475	7,206,853	101.1	90.1
Washington	6,724,540	6,133,267	6,884,715	6,028,786	102.4	89.7
West Virginia	1,852,994	1,684,092	1,827,500	1,657,324	98.6	89.4
Wisconsin	5,686,986	5,379,630	5,739,138	5,309,495	100.9	93.4
Wyoming	563,626	511,595	546,002	503,281	96.9	89.3

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

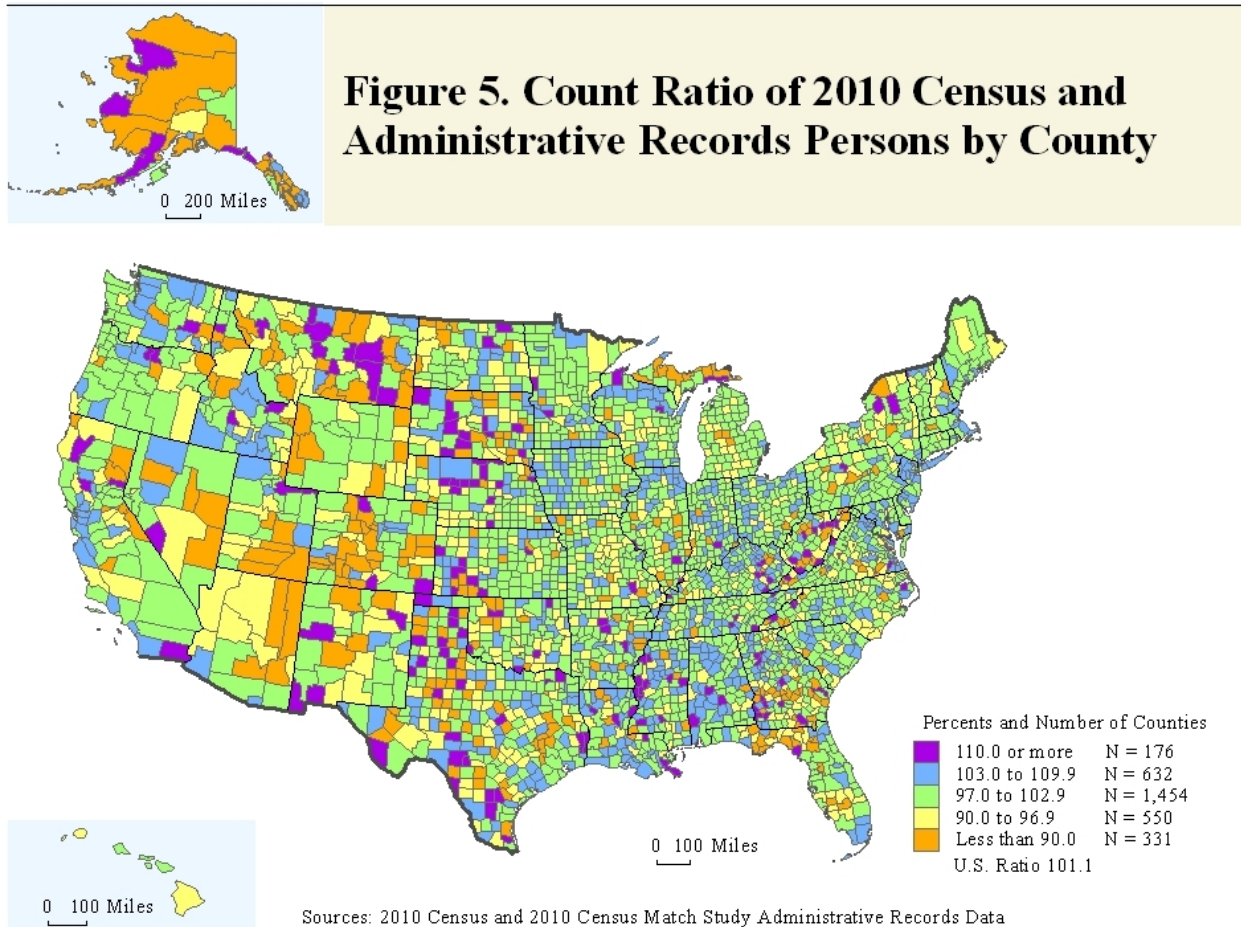
The match ratios ranged from 83.4 percent to 94.1 percent across states. Thirty states had match ratios below 90 percent and twenty-one states had match ratios at 90 percent or above. The states with the highest match ratios were Vermont (94.1 percent), North Dakota (94.1 percent), Iowa (93.6 percent), Wisconsin (93.4 percent), and Maine (93.3 percent). The states with the lowest match ratios were Nevada (83.4 percent), Arizona (84.0 percent), California (84.8 percent), New Mexico (85.0 percent), and the District of Columbia (85.0 percent). Of the ten states that had the highest percentages of 2010 Census and administrative records that matched, six were in the Midwest and four were in the Northeast. Of the ten states that had the lowest percentages of 2010 Census and administrative records that matched, six were in the West, three were in the South, and one was in the Northeast. These results are consistent with the region results, where the Midwest and Northeast had higher match ratios than the South and West.

County

For administrative records, the universe for this sub-section on counties is persons that had information on county of residence. This is slightly lower than the total number of people with PIKs in administrative records because some data sources provided state but not sub-state geographic information. Therefore, there are about 46,000 fewer persons in administrative records represented in this section relative to other sub-sections within the person count and match section.

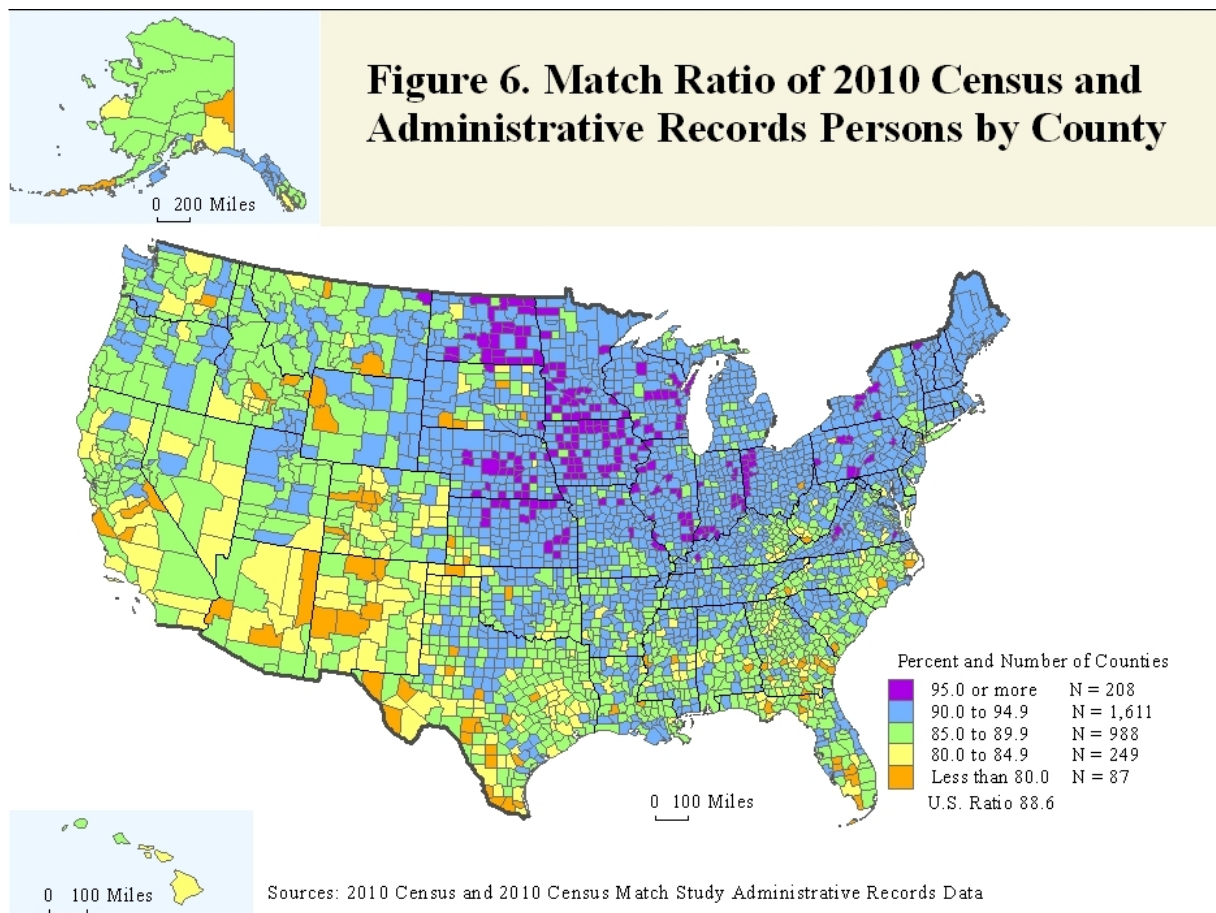
Figure 5 shows person count ratios for the 2010 Census and administrative records by county. Green indicates counties with a count ratio that is closer to 100 percent, yellow and orange indicate low count ratios, and blue and purple represent high count ratios. This map is different from the patterns observed with address county count ratios, where there was a discernible regional and state pattern. This map shows that there were 1,454 counties, almost half of all counties, that had a count ratio close to 100.0 percent and they were distributed relatively evenly across the United States. This is consistent with regional patterns where all region count ratios were similar to the United States count ratio of 101.0 percent.

The person count ratio range across counties was 48.6 percent to 355.2 percent. Of the ten counties with the lowest person count ratios, seven were in the West and three were in the South. Four of the counties in the West were in Colorado: Crowley (48.6 percent), San Juan (58.1 percent), Broomfield (63.4 percent), and Grand (66.2 percent). Three of the counties were in Alaska: Aleutians West (52.3 percent), Wrangell (57.5 percent), and North Slope (58.0 percent). The three counties that were in the South were all in Virginia—Radford (62.5 percent), Lexington (63.8 percent), and Williamsburg (66.5 percent).



Of the ten counties with the highest person count ratios, five were in the West, three in the Midwest, and two in the South. The five counties in the West were Bristol Bay, Alaska (355.2 percent); Lake and Peninsula, Alaska (335.5 percent); Kalawao, Hawaii (332.2 percent); Gilliam, Oregon (266.2 percent); and Sierra, California (209.8 percent). The three counties in the Midwest were Lane, Kansas (251.0 percent); Blaine, Nebraska (216.1 percent); and Hardin, Illinois (185.5 percent). The two counties in the South were McMullen, Texas (325.2 percent) and Roberts, Texas (199.9 percent).

Figure 6 shows 2010 Census and administrative records match ratios by county. Purple represents the counties with highest percent match, followed by blue. Green and yellow represent counties with mid-range match ratios, while orange represents low match ratios.



The person match ratio ranged from 59.4 percent to 97.1 percent across all counties. All states in the Midwest had counties with match ratios of 95.0 percent or above. Three states in the Northeast had counties with match ratios of 95.0 percent or above—Pennsylvania, New York, and Vermont. In the South, two states had counties with match ratios of 95 percent or above—Kentucky and Virginia. The only state in the West that had a county with a match ratio of 95.0 percent or above was Montana. The Midwestern and Northeastern states also had many counties that had match ratios between 90.0 percent and 94.9 percent. Many Southern states also had counties in this range, but less so compared to the Midwest and Northeast. All states in the West, except for Arizona and Hawaii, had at least one county that had a match ratio above 90.0 percent. The majority of counties that had match ratios below 80.0 percent were located in the West and South.

Of the ten counties that had the highest match ratios, nine were in the Midwest and one was in the Northeast. Two were in North Dakota: Foster (97.1 percent) and Emmons (96.6 percent). Two were in Kansas: Republic (96.9 percent) and Marshall (96.6 percent). Two were in Nebraska: Boone (96.9 percent) and Hooker (96.6 percent). Two were in Minnesota: Brown (96.8 percent) and Pope (96.6 percent). The remaining Midwestern county was Carroll, Iowa (96.7 percent). The county in the Northeast was Elk, Pennsylvania (96.8 percent).

Of the counties that had the lowest match ratios, eight of them were in the South. Four were in Texas: Garza (59.4 percent), Concho (60.2 percent), La Salle (63.4 percent), and Reeves (68.3 percent). Two were in Georgia: Stewart (68.4 percent) and Telfair (71.5 percent). The other two counties in the South were in Glades, Florida (70.4 percent) and Issaquena, Mississippi (71.1 percent). Kalawao, Hawaii (70.0 percent) and Shannon, South Dakota (70.3 percent) also were among the ten counties with the lowest match ratios.

The upper and lower bounds of the address count ratio range (12.1 percent to 208.2 percent) were considerably lower than the person count ratio range. The lower bound of the address match ratio range (8.0 percent) was sizably lower than the lower bound for the person match ratio range. Further research should investigate these differences.

Federal and Commercial Data

Table 10 shows person count and match ratios for the 2010 Census and federal and commercial data. In contrast to the federal and commercial address results, federal data had a higher number of persons and higher 2010 Census count and match ratios relative to commercial data. There were 302.2 million persons in the federal administrative records data and 222.0 million persons in the commercial data. The corresponding 2010 Census count ratios were 97.9 percent for federal data and 71.9 percent for commercial data. The match ratio for federal data was 87.4 percent compared to 64.6 percent for commercial data.

Table 10. 2010 Census and Federal and Commercial Administrative Records Person Count and Match Numbers and Ratios

Data Type	Administrative Records Person Count	2010 Census and Administrative Records Person Count Ratio	2010 Census and Administrative Records Person Match Ratio
Commercial	222,021,125	71.9	64.6
Federal	302,191,874	97.9	87.4
In both Commercial and Federal	211,998,674	68.7	63.4

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

There were 212.0 million persons that were in both federal and commercial data. There were a large number of persons that were only found in either commercial data or federal data. However, there were substantially more persons that were only in federal data. There were 10.0 million validated persons that were in commercial data but not in federal data. There were 90.2 million persons that were in federal data but not in commercial data.

Type of Enumeration Area

Table 11 shows 2010 Census and administrative records count and match ratios by TEA.

Table 11. 2010 Census and Administrative Records Person Count and Match Numbers and Ratios by Type of Enumeration Area

Type of Enumeration Area	2010 Census Person Count	Administrative Records Person Count	2010 Census and Administrative Records Person Match	2010 Census and Administrative Records Person Count Ratio	2010 Census and Administrative Records Person Match Ratio
Total	308,745,538	312,214,325	273,643,411	101.1	88.6
Mailout/Mailback	284,908,805	285,001,805	252,750,046	100.0	88.7
Military	922,712	869,278	797,116	94.2	86.4
Remote Alaska	60,261	55,291	51,203	91.8	85.0
Remote Update Enumerate	6,411	5,595	4,605	87.3	71.8
Update Enumerate	2,103,424	2,004,466	1,713,349	95.3	81.5
Update/Leave	15,636,992	14,834,417	13,936,170	94.9	89.1
Urban Update/Leave	5,106,933	4,820,539	4,390,922	94.4	86.0
No TEA	0	4,622,934	0	-	-

Note: A "--" indicates a ratio where the denominator was 0.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

TEA count ratios for persons were higher than for addresses, and the range across TEAs showed less variation. This was to be expected given that TEA is defined by address characteristics, thus it is less likely to affect person counts. The range of TEA count ratios for addresses was 20.0 percent to 200.5 percent, while the TEA count ratio range for persons was 87.3 percent to 100.0 percent. Mailout/Mailback had a count ratio equal to 100.0 percent. There were 284.9 million people in the Mailout/Mailback TEA in the 2010 Census and 285.0 million in administrative records. Update Enumerate had the next highest count ratio (95.3 percent), followed by Update/Leave (94.9 percent), Urban Update/Leave (94.4 percent), and Military (94.2 percent) all of which had count ratios of about 95.0 percent. Remote Alaska had a slightly lower count ratio at 91.8 percent and Remote Update Enumerate had the lowest count ratio (87.3 percent).

The person match ratio also varied less than the address match ratio. The TEA address match ratio ranged from 16.8 percent to 94.6 percent, while the TEA person match ratio ranged from 71.8 percent to 89.1 percent. All TEAs except Remote Update Enumerate had a match ratio above 80.0 percent. Except for Mailout/Mailback and Military TEAs, the person match ratios were higher than the corresponding address match ratios. Update/Leave had the highest match ratio (89.1 percent), followed by Mailout/Mailback (88.7 percent), Military (86.4 percent), Urban Update/Leave (86.0 percent), and Remote Alaska (85.0 percent). The match ratio for Update Enumerate was slightly lower (81.5 percent), and Remote Update Enumerate had the lowest match ratio (71.8 percent).

Demographic Characteristics, Mode, and Proxy

Table 12 shows the number and percentage of PIKs in the 2010 Census and the 2010 Census/administrative records match by demographic characteristics, mode, and proxy.

Table 12. 2010 Census and Administrative Records Person Match by Demographic Characteristics, Mode, and Proxy

Demographic Characteristics, Mode, and Proxy	2010 Census Person Count	2010 Census Persons with a PIK						2010 Census Persons without a PIK	
		Total		2010 Census with a PIK not in Administrative Records		2010 Census and Administrative Records Person Match			
		Number	Number	Percent	Number	Percent	Number	Percent	Number
Total Population	308,745,538	279,179,329	90.4	5,535,918	1.8	273,643,411	88.6	29,566,209	9.6
Hispanic or Latino Origin									
Hispanic	50,477,594	40,554,012	80.3	1,602,206	3.2	38,951,806	77.2	9,923,582	19.7
Not Hispanic	258,267,944	238,625,317	92.4	3,933,712	1.6	234,691,605	90.9	19,642,627	7.6
Race									
White Alone	223,553,254	206,571,803	92.4	3,404,942	1.5	203,166,861	90.9	16,981,451	7.6
Black Alone	38,929,315	34,328,279	88.2	796,386	2.0	33,531,893	86.1	4,601,036	11.8
American Indian or Alaska Native Alone	2,932,370	2,542,640	86.7	45,712	1.6	2,496,928	85.2	389,730	13.3
Asian Alone	14,674,336	12,974,148	88.4	318,390	2.2	12,655,758	86.2	1,700,188	11.6
Native Hawaiian or Other Pacific Islander Alone	540,064	453,090	83.9	15,834	2.9	437,256	81.0	86,974	16.1
Some Other Race Alone	19,107,368	14,232,873	74.5	649,901	3.4	13,582,972	71.1	4,874,495	25.5
Two or More Races	9,008,831	8,076,496	89.7	304,753	3.4	7,771,743	86.3	932,335	10.3
Age									
0-2	12,019,146	10,776,958	89.7	1,337,667	11.1	9,439,291	78.5	1,242,188	10.3
3-17	62,162,321	56,554,181	91.0	2,647,192	4.3	53,906,989	86.7	5,608,140	9.0
18-24	30,646,519	26,147,233	85.3	478,323	1.6	25,668,910	83.8	4,499,286	14.7
25-44	82,123,330	72,072,154	87.8	670,314	0.8	71,401,840	86.9	10,051,176	12.2
45-64	81,499,596	75,765,796	93.0	284,406	0.3	75,481,390	92.6	5,733,800	7.0
65-74	21,727,578	20,502,704	94.4	59,034	0.3	20,443,670	94.1	1,224,874	5.6
75 and older	18,567,048	17,360,303	93.5	58,982	0.3	17,301,321	93.2	1,206,745	6.5
Sex									
Male	151,775,099	136,105,431	89.7	3,020,094	2.0	133,085,337	87.7	15,669,668	10.3
Female	156,970,439	143,073,898	91.1	2,515,824	1.6	140,558,074	89.5	13,896,541	8.9
Mode									
Nonresponse Followup	60,432,209	49,285,340	81.6	1,239,354	2.1	48,045,986	79.5	11,146,869	18.4
Mailout/Mailback	205,816,623	198,977,997	96.7	2,891,481	1.4	196,086,516	95.3	6,838,626	3.3
Other	42,496,706	30,915,992	72.7	1,405,083	3.3	29,510,909	69.4	11,580,714	27.3
Proxy									
Not by Proxy	295,163,226	274,587,574	93.0	5,463,417	1.9	269,124,157	91.2	20,575,652	7.0
By Proxy	13,582,312	4,591,755	33.8	72,501	0.5	4,519,254	33.3	8,990,557	66.2

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Note that the characteristic, mode, count imputation, and proxy data in Table 12 is from the 2010 Census, thus the Hispanic origin and race analysis is not based on matched Hispanic origin and race responses in the 2010 Census and administrative records. Matched demographic response data will be evaluated in section 5.4 of this report.

A higher percentage of the non-Hispanic population was PIKed in the 2010 Census relative to the Hispanic population. The non-Hispanic population also had a higher percentage that was in both the 2010 Census and administrative records relative to the Hispanic population. Of the 258.3 million non-Hispanics in the 2010 Census, 238.6 million or 92.4 percent were PIKed, and 234.7 million or 90.9 percent were in both the 2010 Census and administrative records.

While these results were lower for the Hispanic population, administrative records covered a substantial proportion of the Hispanic population in the 2010 Census. Of the 50.5 million Hispanics in the 2010 Census, 40.6 million or 80.3 percent were PIKed, and 39.0 million or 77.2 percent were in both the 2010 Census and administrative records.

The percentage of persons PIKed in the 2010 Census by race group ranged from 74.5 percent to 92.4 percent. The percentage of persons in the 2010 Census by race group who were also in administrative records was similar to, yet slightly lower than, the percentage PIKed in the 2010 Census, 71.1 percent to 90.9 percent.

The White alone population had the highest percentage PIKed in the 2010 Census and the highest percentage in both the 2010 Census and administrative records relative to all other race groups. Of the 223.6 million persons classified as White alone in the 2010 Census, 206.6 million or 92.4 percent were PIKed, and 203.2 million or 90.9 percent were in the 2010 Census and administrative records. The Two or More Races population had the second highest percentage PIKed in the 2010 Census and the second highest percentage also in administrative records. Of the 9.0 million persons classified as Two or More Races, 8.1 million or 89.7 percent were PIKed, and 7.8 million or 86.3 percent were in the 2010 Census and administrative records.

The Asian alone population had the third highest percentage PIKed (88.4 percent) in the 2010 Census and the third highest percentage that was in the 2010 Census and administrative records (86.2 percent), followed by the Black alone population, the AIAN alone population, and the NHPI alone population.

The SOR alone population had the lowest percentage (74.5 percent) PIKed in the 2010 Census and the lowest percentage in both the 2010 Census and administrative records (71.1 percent). This lower PIK percentage for the SOR alone population was largely driven by the Hispanic population, as 96.8 percent of those classified as SOR alone in the 2010 Census were of Hispanic origin (Humes et al. 2011).

The percentage of males PIKed in the 2010 Census was slightly lower than the percentage of females. Of the 151.8 million males in the 2010 Census, 136.1 million or 89.7 percent were PIKed, and 133.1 million or 87.7 percent were in both the 2010 Census and administrative records. Of the 157.0 million females in the 2010 Census, 143.1 million or 91.1 percent were PIKed, and 140.6 million or 89.5 percent were in both the 2010 Census and administrative records.

For age groups, the percentage PIKed in the 2010 Census ranged from 85.3 percent to 94.4 percent. The proportions in the 2010 Census and administrative records were slightly lower and ranged from 78.5 percent to 94.1 percent. Older age groups had higher proportions that were PIKed and in the 2010 Census and administrative records relative to younger age groups. The age group 65 to 74 had the highest percentage PIKed (94.4 percent) and in administrative records (94.1 percent). Of the 21.7 million persons aged 65 to 74, 20.5 million were PIKed and about the same number were found in administrative records. The age group of 75 and older had the second highest percentage PIKed (93.5 percent) and the second highest percentage in the 2010 Census and administrative records (93.2 percent). The age group 45 to 64 had the next highest percentages that were PIKed (93.0 percent) and also in administrative records (92.6 percent), followed by the age group 3 to 17 (91.0 percent and 86.7 percent). The age group 18 to 24 had the lowest percentage PIKed (85.3 percent) and the second lowest percentage in administrative records (83.8 percent).

For the age group 0 to 2, 89.7 percent were PIKed, but this age group had the lowest proportion in both the 2010 Census and administrative records at 78.5 percent. More than 11 percent of this age group was in the 2010 Census with a PIK but not in administrative records. This may be due in part to the tax filing issues discussed at the beginning of the person section. Those aged 3 to 17 were also less likely than other groups to be in both the 2010 Census and administrative records, where 4.3 percent of this age group was in the 2010 Census with a PIK but not in administrative records as compared to 1.6 percent or less for the age group 18 to 24.

A higher percentage of persons in the 2010 Census that lived in households that responded by mail were PIKed, and these persons were also more likely to be in administrative records compared to NRFU and other modes. Of the 205.8 million persons that were in households that responded by mail, 96.7 percent were PIKed and 95.3 percent were in administrative records. Of the 13.6 million proxy responses, a low percent were PIKed and were also in administrative records, about 33 percent. Of the 60.4 million persons in the 2010 Census that responded via NRFU, 49.3 million were PIKed and 48.0 million or 79.5 percent were in administrative records.

5.3 Person-Address Pair Count and Match

Nation

This section assesses administrative data relative to the 2010 Census after the best address model has been applied to select the best address for Census Day in the administrative data.²² As discussed above, the 2010 Census also has the same PIK at multiple addresses, and these duplicate person-address pairs in the 2010 Census are included in the following analysis.

Figure 7 shows the number and match of 2010 Census and administrative records person-address pairs.²³ All persons in the 2010 Census were associated with an address, thus all 2010 Census person count and PIK numbers discussed in the person count and match section are the same in Figure 7. For instance, there were 308.7 million people in the 2010 Census with an address.

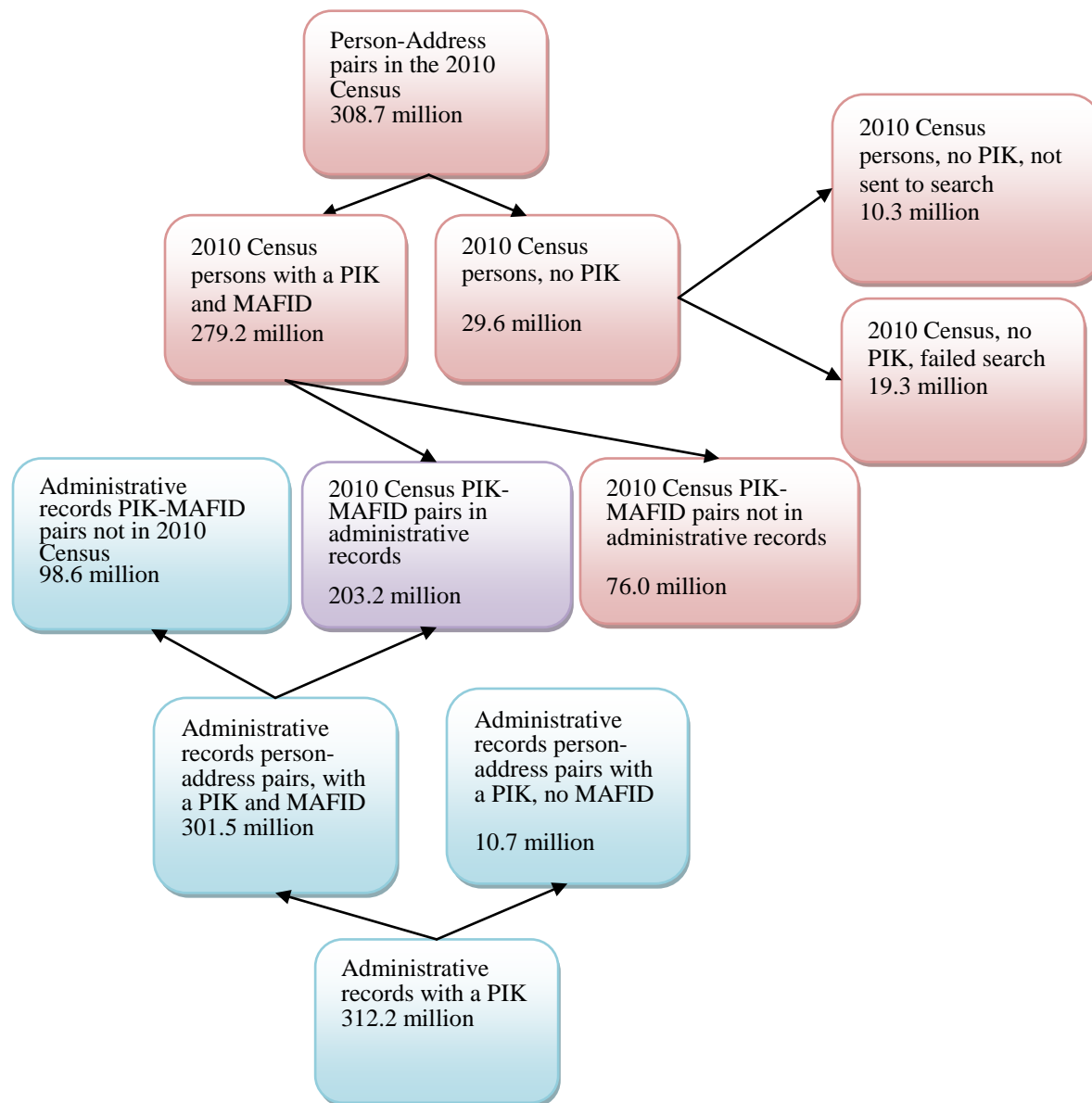
As noted in the person count and match section, there were 312.2 million persons in administrative records that had a PIK and were alive on Census Day. Of those, 301.5 million PIKed persons had one or more MAFIDs, and 10.7 million PIKed persons did not have a MAFID. Before we applied the best address model, there were 216.2 million 2010 Census person-address pairs that matched to administrative records. Of the 308.7 million persons in the 2010 Census, 70.0 percent matched to administrative records person-address pairs. Of the 279.2 million person-address pairs in the 2010 Census that had a PIK, 77.4 percent matched to administrative records person-address pairs. After applying the best address model to administrative records with multiple MAFIDs, there were 203.2 million 2010 Census person-address pairs that matched to administrative records. Of the 308.7 million persons in the 2010 Census, 65.8 percent matched to administrative records person-address pairs. Of the 279.2 million persons in the 2010 Census that had a PIK, 72.8 percent matched to administrative records person-address pairs.

There were 76.0 million 2010 Census person-address pairs with a PIK and MAFID that did not match to administrative records. There were 98.6 million administrative records person-address pairs with a PIK and MAFID that did not match to the 2010 Census.

²² The best address model was applied to the PIKs in administrative records with two or more MAFIDs. Among those PIKs with a MAFID, about 152.8 million PIKs (50.7 percent) had exactly one unique MAFID. Of those PIKs with multiple associated MAFIDs, 75.4 million (25.0 percent) had two MAFIDs, and 39.7 million (13.2 percent) had three MAFIDs. Another 19.2 million PIKs (6.4 percent) had four unique MAFIDs in the administrative records, and 8.5 million PIKs (2.8 percent) had five MAFIDs. The remaining 5.9 million PIKs with MAFIDs in the administrative records had six or more unique MAFIDs associated with them.

²³ The 2010 Census included duplicate person-address pairs whereas the administrative records contained unique person-address pairs. This resulted in instances where a single administrative record person-address pair matched to multiple census record person-address pairs. Therefore, the sum of the count for administrative records PIK-MAFID pairs not in 2010 Census (98.6 million) and the count for 2010 Census PIK-MAFID pairs in administrative records (203.2 million) does not equal the number of administrative records PIK-MAFID pairs (301.5 million).

Figure 7. Count and Match of 2010 Census and Administrative Records Person-Address Pairs



Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Region

Table 13 shows the 2010 Census person-address count, administrative records person-address count, the number of 2010 Census and administrative records person-address pairs that matched, and the 2010 Census and administrative records person-address count and match ratios by region.

Table 13. 2010 Census and Administrative Records Person-Address Count and Match Numbers and Ratios by Region

Region	2010 Census Person-Address Count	Administrative Records Person- Address Count	2010 Census and Administrative Records Person-Address Match	2010 Census and Administrative Records Person- Address Count Ratio	2010 Census and Administrative Records Person- Address Match Ratio
Total	308,745,538	301,516,209	203,157,426	97.7	65.8
Northeast	55,317,240	53,973,110	36,432,719	97.6	65.9
Midwest	66,927,001	66,094,806	47,943,123	98.8	71.6
South	114,555,744	111,709,332	73,198,676	97.5	63.9
West	71,945,553	69,738,961	45,582,908	96.9	63.4

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The person-address count ratio for the United States was 97.7 percent. Across all regions, the person-address count ratio was close to the national count ratio. The person-address count ratio was highest for the Midwest at 98.8 percent, followed by the Northeast (97.6 percent), South (97.5 percent), and West (96.9 percent).

The match ratio for the United States was substantially lower than the count ratio (65.8 percent), and this was reflected across the regions. The person-address match ratio ranking among regions was the same as for the count ratios, where the Midwest had the highest match ratio (71.6 percent), followed by the Northeast (65.9 percent), South (63.9 percent), and West (63.4 percent).

State

Table 14 shows the 2010 Census person-address count, administrative records person-address count, the number of 2010 Census and administrative records person-address pairs that matched, and the 2010 Census and administrative records person-address count and match ratios by state.

The five states that had the highest person-address count ratios were Maryland (100.4 percent), Ohio (100.1 percent), Illinois (99.9 percent), Delaware (99.3 percent), and New Jersey (99.3 percent). Of the ten states with the highest count ratios, five were in the Midwest, three in the South, and one in the West.

The five states with the lowest count ratios were Alaska (84.1 percent), Wyoming (85.2 percent), New Mexico (87.8 percent), Montana (89.2 percent), and West Virginia (91.2 percent). Of the ten states with the lowest count ratios, seven were in the West, one in the South, one in the Midwest, and one in the Northeast. These results are consistent with the region person-address count ratios, where the Midwest had the highest count ratios and the West had the lowest.

Table 14. 2010 Census and Administrative Records Person-Address Count and Match Numbers and Ratios by State

State	2010 Census Person-Address Count	Administrative Records Person- Address Count	2010 Census and Administrative Records Person- Address Match	2010 Census and Administrative Records Person- Address Count Ratio	2010 Census and Administrative Records Person- Address Match Ratio
Total	308,745,538	301,516,209	203,157,426	97.7	65.8
Alabama	4,779,736	4,680,999	2,981,411	97.9	62.4
Alaska	710,231	597,613	340,527	84.1	47.9
Arizona	6,392,017	5,882,725	3,702,602	92.0	57.9
Arkansas	2,915,918	2,769,483	1,795,591	95.0	61.6
California	37,253,956	36,895,430	23,858,501	99.0	64.0
Colorado	5,029,196	4,864,921	3,277,496	96.7	65.2
Connecticut	3,574,097	3,492,906	2,523,428	97.7	70.6
Delaware	897,934	891,639	623,461	99.3	69.4
District of Columbia	601,723	591,770	342,003	98.3	56.8
Florida	18,801,310	18,571,203	12,167,579	98.8	64.7
Georgia	9,687,653	9,548,384	5,996,844	98.6	61.9
Hawaii	1,360,301	1,253,669	741,802	92.2	54.5
Idaho	1,567,582	1,448,474	1,007,470	92.4	64.3
Illinois	12,830,632	12,822,700	8,630,674	99.9	67.3
Indiana	6,483,802	6,416,121	4,675,947	99.0	72.1
Iowa	3,046,355	2,977,126	2,266,850	97.7	74.4
Kansas	2,853,118	2,792,230	2,034,442	97.9	71.3
Kentucky	4,339,367	4,218,816	2,851,115	97.2	65.7
Louisiana	4,533,372	4,411,361	2,779,649	97.3	61.3
Maine	1,328,361	1,278,617	862,986	96.3	65.0
Maryland	5,773,552	5,794,145	4,121,327	100.4	71.4
Massachusetts	6,547,629	6,453,301	4,528,654	98.6	69.2
Michigan	9,883,640	9,667,350	7,092,248	97.8	71.8
Minnesota	5,303,925	5,245,597	4,014,818	98.9	75.7
Mississippi	2,967,297	2,857,348	1,722,241	96.3	58.0
Missouri	5,988,927	5,830,474	4,120,999	97.4	68.8
Montana	989,415	882,079	569,270	89.2	57.5
Nebraska	1,826,341	1,780,571	1,323,040	97.5	72.4
Nevada	2,700,551	2,654,172	1,616,682	98.3	59.9
New Hampshire	1,316,470	1,286,020	927,007	97.7	70.4
New Jersey	8,791,894	8,727,028	5,963,720	99.3	67.8
New Mexico	2,059,179	1,807,812	1,056,957	87.8	51.3
New York	19,378,102	18,666,689	11,472,664	96.3	59.2
North Carolina	9,535,483	9,169,433	6,191,068	96.2	64.9
North Dakota	672,591	623,567	448,212	92.7	66.6
Ohio	11,536,504	11,552,963	8,518,977	100.1	73.8
Oklahoma	3,751,351	3,577,427	2,219,125	95.4	59.2
Oregon	3,831,074	3,716,295	2,610,007	97.0	68.1
Pennsylvania	12,702,379	12,482,815	9,075,510	98.3	71.4
Rhode Island	1,052,567	1,005,285	692,881	95.5	65.8
South Carolina	4,625,364	4,475,235	3,022,905	96.8	65.4
South Dakota	814,180	766,213	534,715	94.1	65.7
Tennessee	6,346,105	6,290,515	4,319,859	99.1	68.1
Texas	25,145,561	24,293,996	15,479,039	96.6	61.6
Utah	2,763,885	2,693,874	1,884,028	97.5	68.2
Vermont	625,741	580,449	385,869	92.8	61.7
Virginia	8,001,024	7,877,584	5,648,319	98.5	70.6
Washington	6,724,540	6,561,481	4,598,158	97.6	68.4
West Virginia	1,852,994	1,689,994	937,140	91.2	50.6
Wisconsin	5,686,986	5,619,894	4,282,201	98.8	75.3
Wyoming	563,626	480,416	319,408	85.2	56.7

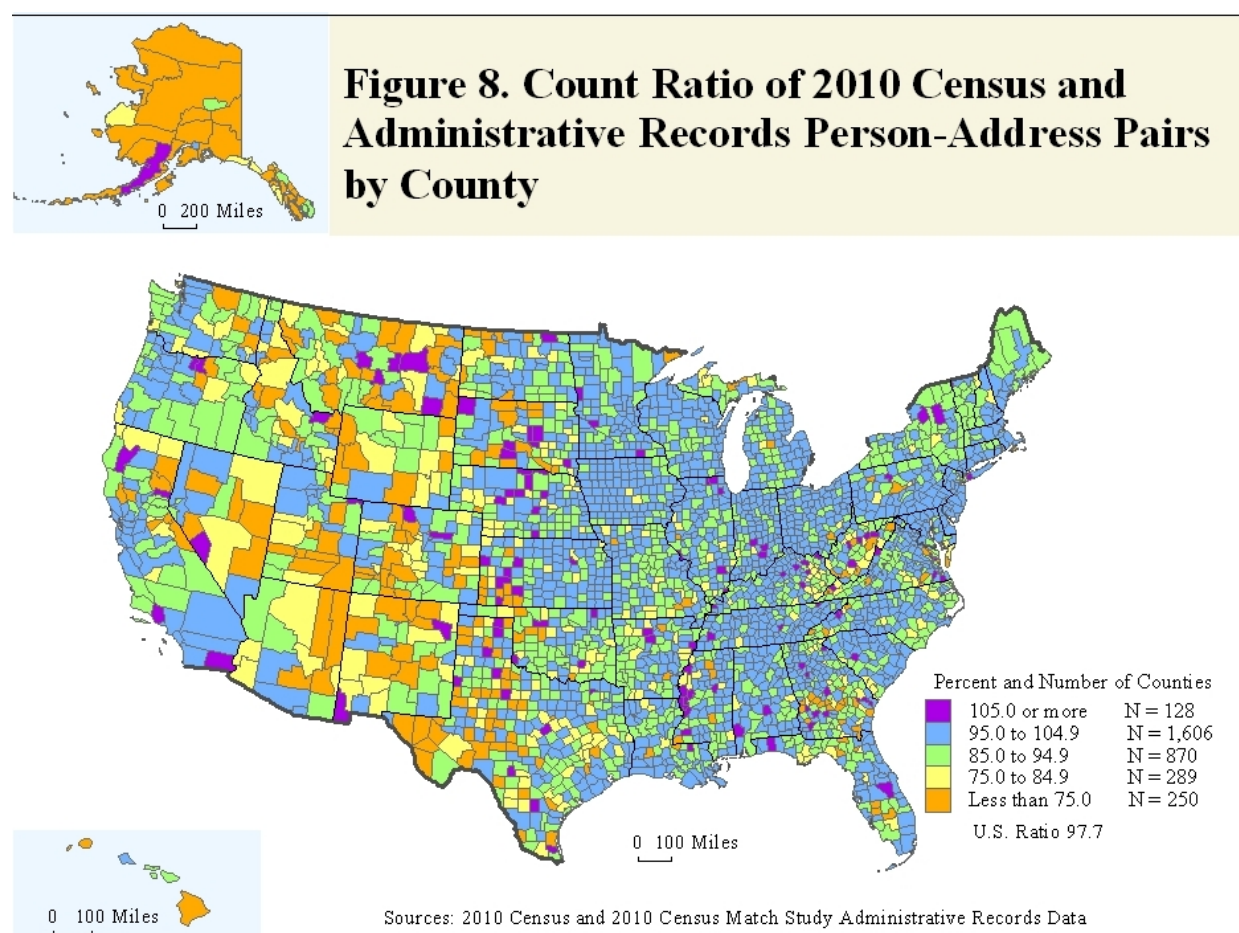
Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The five states with the highest person-address match ratios were all in the Midwest: Minnesota (75.7 percent), Wisconsin (75.3 percent), Iowa (74.4 percent), Ohio (73.8 percent), and Nebraska (72.4 percent). Of the ten states with the highest match ratios, eight were in the Midwest, one in the South, and one in the Northeast.

The five states with the lowest person-address match ratios were Alaska (47.9 percent), West Virginia (50.6 percent), New Mexico (51.3 percent), Hawaii (54.5 percent), and Wyoming (56.7 percent). Of the ten states with the lowest match ratios, six were in the West and four were in the South. Consistent with address results, states with low person-address count and match ratios tended to have fewer person-address pairs in Mailout/Mailback TEAs relative to states that had high count and match ratios.

County

Figure 8 shows the person-address count ratios by county. Blue indicates counties with a count ratio close to 100 percent.



Many states in the Midwest had counties with count ratios close to 100 percent, such as Iowa, Wisconsin, Illinois, Indiana, and Ohio. In the Northeast, a few states had counties with count

ratios around 100 percent such as Pennsylvania and New Jersey. In the South, states that had counties with count ratios around 100 percent included Maryland, Tennessee, Kentucky, Delaware, and Alabama. Many states in the West had counties with low count ratios relative to the Midwest and South. These states include Wyoming, Arizona, New Mexico, and Alaska.

Figure 9 displays person-address match ratios by county. Purple and blue indicate counties with higher match ratios, while yellow and orange represent low match ratios. States in the Midwest that had counties with high match ratios include Iowa, Minnesota, Wisconsin, Ohio, and Indiana. In the South, states such as Virginia, Maryland, and Tennessee had counties with high match ratios. In the Northeast, states with high match ratios included Pennsylvania and New Jersey. Many states in the West and South had a number of counties with low match ratios.

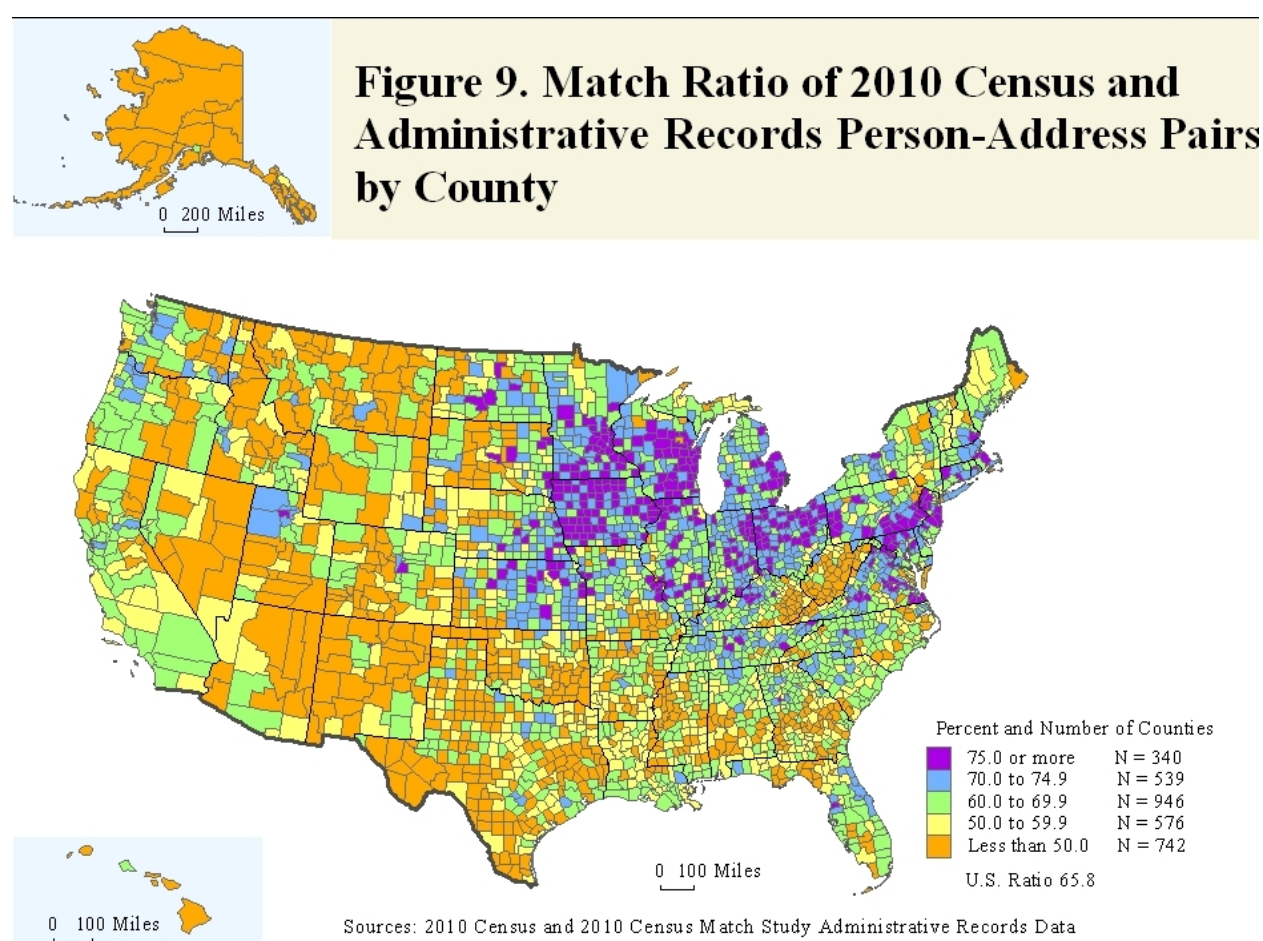


Table 15 shows count ratios, match ratios, and TEA by county. Of the ten counties in the United States that had the lowest count ratios, eight were in Alaska: North Slope (17.1 percent), Aleutians West (19.3 percent), Wrangell (21.7 percent), Bethel (25.2 percent), Nome (26.0 percent), Haines (26.3 percent), Petersburg (27.3 percent), and Yukon-Koyukuk (29.2 percent). One county in Wyoming and one county in South Dakota were also among the ten counties with the lowest count ratios, Teton (29.2 percent) and Todd (31.9 percent), respectively.

Table 15. 2010 Census and Administrative Records Person-Address Count Ratio, Match Ratio, and Type of Enumeration Area for the Ten Counties with the Lowest and Highest Ratios

County	Ratio	Type of Enumeration Area						
		Mailout / Mailback	Military	Remote Alaska	Remote Update Enumerate	Update Enumerate	Update / Leave	Urban Update / Leave
Lowest Count Ratios								
North Slope, Alaska	17.1	0.0	0.0	55.3	0.0	0.0	44.7	0.0
Aleutians West, Alaska	19.3	0.0	0.0	21.3	0.0	0.0	78.7	0.0
Wrangell, Alaska	21.7	0.0	0.0	0.0	7.0	0.0	93.0	0.0
Bethel, Alaska	25.2	0.0	0.0	64.3	0.0	0.0	35.7	0.0
Nome, Alaska	26.0	0.0	0.0	61.1	0.0	0.0	38.9	0.0
Haines, Alaska	26.3	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Petersburg, Alaska	27.3	0.0	0.0	0.0	7.4	0.0	92.6	0.0
Teton, Wyoming	29.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Yukon-Koyukuk, Alaska	29.2	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Todd, South Dakota	31.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Highest Count Ratios								
Kalawao, Hawaii	330.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Bristol Bay, Alaska	323.6	0.0	0.0	100.0	0.0	0.0	0.0	0.0
McMullen, Texas	300.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Lake and Peninsula, Alaska	296.4	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Gilliam, Oregon	248.0	15.2	0.0	0.0	0.0	0.0	84.8	0.0
Lane, Kansas	232.0	17.8	0.0	0.0	0.0	0.0	82.2	0.0
Blaine, Nebraska	204.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Roberts, Texas	191.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Sierra, California	187.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Hardin, Illinois	170.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Lowest Match Ratios								
Aleutians East, Alaska	0.9	0.0	0.0	65.5	0.0	0.0	34.5	0.0
Aleutians West, Alaska	1.5	0.0	0.0	21.3	0.0	0.0	78.7	0.0
Kalawao, Hawaii	2.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Yukon-Koyukuk, Alaska	2.4	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Shannon, South Dakota	2.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Yakutat, Alaska	3.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Northwest Arctic, Alaska	3.0	0.0	0.0	57.5	0.0	0.0	42.5	0.0
Dillingham, Alaska	3.1	0.0	0.0	51.9	0.0	0.0	48.1	0.0
Todd, South Dakota	3.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Nome, Alaska	3.7	0.0	0.0	61.1	0.0	0.0	38.9	0.0
Highest Match Ratios								
Poquoson, Virginia	85.1	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Medina, Ohio	83.7	99.2	0.0	0.0	0.0	0.0	0.8	0.0
Ozaukee, Wisconsin	83.5	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Monroe, Illinois	83.5	95.6	0.0	0.0	0.0	0.0	4.4	0.0
Anoka, Minnesota	83.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Wood, Wisconsin	83.0	97.9	0.0	0.0	0.0	0.2	1.9	0.0
Washington, Wisconsin	82.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Scott, Minnesota	82.8	99.9	0.0	0.0	0.0	0.0	0.1	0.0
Waukesha, Wisconsin	82.8	99.8	0.0	0.0	0.0	0.0	0.2	0.0
Washington, Minnesota	82.6	100.0	0.0	0.0	0.0	0.0	0.0	0.0

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Kalawao, Hawaii (330.0 percent) had the highest county count ratio, followed by Bristol Bay, Alaska (323.6 percent). Among the counties with the highest count ratios, three additional counties were in the West: Lake and Peninsula, Alaska (296.4 percent); Gilliam, Oregon (248.0 percent); and Sierra, California (187.0 percent). Two were in the South in Texas: McMullen (300.7 percent) and Roberts (191.0 percent). Three were in the Midwest: Lane, Kansas (232.0 percent); Blaine, Nebraska (204.0 percent); and Hardin, Illinois (170.0 percent).

Seven of the ten counties with the lowest match ratios were in Alaska: Aleutians East (0.9 percent), Aleutians West (1.5 percent), Yukon-Koyukuk (2.4 percent), Yakutat (3.0 percent), Northwest Arctic (3.0 percent), Dillingham (3.1 percent), and Nome (3.7 percent). Two counties were in South Dakota: Shannon (2.4 percent) and Todd (3.4 percent), and one county was in Hawaii: Kalawao (2.2 percent).

Of the ten counties with the highest match ratios, Poquoson, Virginia had the highest at 85.1 percent. The remaining nine counties were in the Midwest. Four of the counties were in Wisconsin: Ozaukee (83.5 percent), Wood (83.0 percent), Washington (82.9 percent), and Waukesha (82.8 percent). Three were in Minnesota: Anoka (83.4 percent), Scott (82.8 percent), and Washington (82.6 percent). One county was in Ohio: Medina (83.7 percent), and one county was in Illinois: Monroe (83.5 percent). For the person-address match ratios, as was observed for addresses, of the ten counties with the lowest and highest match ratios, counties that had more TEAs designated as Mailout/Mailback had higher matches. There was no discernible TEA pattern for count ratios.

Federal and Commercial Data

Table 16 shows count and match ratios for the 2010 Census and federal and commercial data.

Table 16. 2010 Census and Federal and Commercial Administrative Records Person-Address Count and Match Numbers and Ratios

Data Type	Administrative Records Person- Address Count	2010 Census and Administrative Records Person- Address Count Ratio	2010 Census and Administrative Records Person- Address Match Ratio
Commercial	219,466,721	71.1	48.8
Federal	292,328,979	94.7	65.4
In both Commercial and Federal	210,279,491	68.1	48.3

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Federal data had a higher number of person-address pairs and higher 2010 Census count and match ratios relative to commercial data. There were 292.3 million PIKs in federal data with a best address assigned in administrative records, resulting in a 2010 Census count ratio of 94.7 percent. There were 201.9 million 2010 Census records that matched to federal administrative records for a match ratio of 65.4 percent. There were 219.5 million PIKs in commercial data

with a best address assigned in administrative records, and the 2010 Census count ratio was 71.1 percent. There were 150.6 million 2010 Census records that matched commercial data for a match ratio of 48.8 percent.

There were 210.3 million person-address pairs that were found in both federal and commercial data. There were a large number of person-address pairs that were only found in either commercial data or federal data. However, similar to the person results, there were substantially more person-address pairs that were only in federal data relative to commercial data. There were 9.2 million person-address pairs that were in commercial data but not in federal data. There were 82.0 million person-address pairs that were in federal data but not in commercial data.

Type of Enumeration Area

Table 17 shows 2010 Census and administrative records person-address count and match ratios by TEA.

Table 17. 2010 Census and Administrative Records Person-Address Count and Match Numbers and Ratios by Type of Enumeration Area

Type of Enumeration Area	2010 Census Person-Address Count	Administrative Records Person-Address Count	2010 Census and Administrative Records Person-Address Match	2010 Census and Administrative Records Person-Address Count Ratio	2010 Census and Administrative Records Person-Address Match Ratio
Total	308,745,538	301,516,209	203,157,426	97.7	65.8
Mailout/Mailback	284,908,805	280,093,025	191,914,484	98.3	67.4
Military	922,712	619,979	358,116	67.2	38.8
Remote Alaska	60,261	35,019	2,902	58.1	4.8
Remote Update Enumerate	6,411	5,278	1,679	82.3	26.2
Update Enumerate	2,103,424	1,700,836	801,040	80.9	38.1
Update/Leave	15,636,992	12,922,334	7,787,827	82.6	49.8
Urban Update/Leave	5,106,933	4,210,134	2,291,378	82.4	44.9
No TEA	0	1,929,604	0	-	-

Note: A “-” indicates a ratio where the denominator was 0.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The count ratios for TEA ranged from 58.1 percent to 98.3 percent. The Mailout/Mailback TEA had the highest count ratio at 98.3 percent, followed by Update/Leave (82.6 percent), Urban Update/Leave (82.4 percent), Remote Update Enumerate (82.3 percent), Update Enumerate (80.9 percent), Military (67.2 percent), and Remote Alaska (58.1 percent).

The match ratios were considerably lower than the count ratios. The match ratios ranged from 4.8 percent to 67.4 percent. The Mailout/Mailback (67.4 percent), Update/Leave (49.8 percent), and Urban Update/Leave (44.9 percent) TEAs had the highest match ratios; followed by Military (38.8 percent), Update Enumerate (38.1 percent), and Remote Update Enumerate (26.2 percent). Remote Alaska had the lowest match ratio at 4.8 percent.

Demographic Characteristics and Census Operations

Table 18 shows 2010 Census and administrative records person-address match ratios by race, Hispanic origin, age, sex, mode, and proxy.

Table 18. 2010 Census and Administrative Records Person-Address Match by Race, Hispanic Origin, Age, Sex, Mode, and Proxy

Demographic Characteristics, Mode, and Proxy	2010 Census Person-Address Count	2010 Census and Administrative Records Person-Address Match	2010 Census and Administrative Records Person-Address Match Ratio
Total Population	308,745,538	203,157,426	65.8
Hispanic or Latino Origin			
Hispanic	50,477,594	26,854,907	53.2
Not Hispanic	258,267,944	176,302,519	68.3
Race			
White Alone	223,553,254	155,730,544	69.7
Black Alone	38,929,315	21,472,380	55.2
American Indian or Alaska Native Alone	2,932,370	1,360,223	46.4
Asian Alone	14,674,336	9,831,674	67.0
Native Hawaiian or Other Pacific Islander Alone	540,064	287,415	53.2
Some Other Race Alone	19,107,368	9,052,400	47.4
Two or More Races	9,008,831	5,422,790	60.2
Age			
0-2	12,019,146	6,685,410	55.6
3-17	62,162,321	39,928,333	64.2
18-24	30,646,519	14,815,295	48.3
25-44	82,123,330	51,755,207	63.0
45-64	81,499,596	60,092,094	73.7
65-74	21,727,578	16,699,927	76.9
75 and older	18,567,048	13,181,160	71.0
Sex			
Male	151,775,099	97,583,770	64.3
Female	156,970,439	105,573,656	67.3
Mode			
Nonresponse Followup	60,432,209	28,721,088	47.5
Mailout/Mailback	205,816,623	158,248,584	76.9
Other	42,496,706	16,187,754	38.1
Proxy			
Not by Proxy	295,163,226	200,630,386	68.0
By Proxy	13,582,312	2,527,040	18.6

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Similar to the person results, a higher percentage of non-Hispanic person-address pairs in the 2010 Census matched to administrative records relative to Hispanics. Of the 258.3 million non-Hispanics in the 2010 Census, 176.3 million or 68.3 percent matched to administrative records person-address pairs. Of the 50.5 million Hispanics in the 2010 Census, 53.2 percent matched to administrative records person-address pairs.

The match ratio ranged from 46.4 percent to 69.7 percent across race groups. Similar to the person results, the White alone population had the highest percentage of 2010 Census records that matched to administrative records person-address pairs. Of the 223.6 million persons in the 2010 Census that were classified as White alone, 155.7 million or 69.7 percent matched to administrative records person-address pairs. The Asian alone population had the second highest match ratio at 67.0 percent, followed by the Two or More Races population (60.2 percent), the Black alone population (55.2 percent), the NHPI alone population (53.2 percent), and SOR alone population (47.4 percent). The AIAN alone population had the lowest match ratio at 46.4 percent.

The person-address match ratio ranged from 48.3 percent to 76.9 percent across age groups. The person-address results follow the same pattern as the person results for age, where match ratios were higher for the older age groups and lower for younger age groups. The age group 65 to 74 had the highest match ratio (76.9 percent), followed by those aged 45 to 64 (73.7 percent). The age group 18 to 24 had the lowest match ratio at 48.3 percent. The age group 0 to 2 had the second lowest match ratio (55.6 percent).

Consistent with the person results, the match ratios for males and females were similar, and females had a slightly higher match ratio. The match ratio for females was 67.3 percent, and the match ratio for males was 64.3 percent.

Similar to the address and person results, a larger number and percentage of 2010 Census person-address pairs that responded via Mailout/Mailback matched to administrative records compared to NRFU and other modes. Of the 205.8 million persons in the 2010 Census that responded via Mailout/Mailback, 158.2 million or 76.9 percent were in administrative records. Of the 60.4 million 2010 Census person-address pairs in NRFU, 28.7 million or 47.5 percent matched to administrative records.

Similar to but even lower than the person results, a low number and percentage of 2010 Census person-address pairs that had a proxy response were in administrative records. Of the 13.6 million responses in the 2010 Census that were provided via proxy, administrative record person-address pairs matched to 2.5 million or 18.6 percent.

The preceding results indicate that direct replacement of administrative records data would result in variable coverage across states and could produce undercounts for various race, Hispanic origin, and age groups. The 2010 Census Match Study was designed to evaluate the quality and

coverage of administrative records data relative to the 2010 Census. The person-address section, as with the address and person sections, reflect different dimensions of the administrative records data to inform future planning and operational uses.

Occupancy Status

Table 19 shows 2010 Census and administrative records by occupancy status.

Table 19. 2010 Census and Administrative Records by Housing Unit Status

Housing Unit Status	2010 Census Housing Unit Count	2010 Census and Administrative Records Same Housing Unit Status		2010 Census and Administrative Records Different Housing Unit Status	
	Number	Number	Percent	Number	Percent
Total	136,592,084	111,659,541	81.7	24,932,543	18.3
Occupied	116,716,292	96,083,076	82.3	20,633,216	17.7
Vacant	14,988,438	11,404,442	76.1	3,583,996	23.9
Delete	4,887,354	4,172,023	85.4	715,331	14.6

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

There were 136.6 million addresses in the 2010 Census that had an occupancy status of occupied, vacant, or delete. Administrative records can inform whether a housing unit is occupied if there is a person in administrative records that lives at a particular housing unit. Administrative records can indicate whether a unit is not occupied if there is no person at that address in federal or commercial files. Units not occupied in administrative records may have either vacant or delete status in the 2010 Census.

Of the 136.6 million 2010 Census addresses, administrative record person-addresses pairs had the same housing unit status for 111.7 million or 81.7 percent of addresses. Of the 116.7 million housing units that were designated as occupied in the 2010 Census, administrative records indicated that 96.1 million or 82.3 percent of these addresses were occupied. Administrative records indicated that the remaining 20.6 million addresses were not occupied.

The 2010 Census had 15.0 million addresses that were designated as vacant. Administrative records indicated that 11.4 million or 76.1 percent of these 15.0 million addresses were not occupied, but that 3.6 million or 23.9 percent were occupied. In the 2010 Census, there were 4.9 million addresses that were designated as deletes. Administrative records found that 4.2 million addresses or 85.4 percent were not occupied and approximately 715,000 addresses or 14.6 percent were occupied in administrative records.

Table 20 shows 2010 Census and administrative records housing unit status by mode.

Of the occupied housing units in the 2010 Census (116.7 million), 82.3 million responded via Mailout/Mailback. For 72.1 million or 87.6 percent of these addresses, administrative records

also found the address to be occupied and 10.2 million or 12.4 percent were vacant. This percentage is lower for both the other and Nonresponse Followup mode categories. Of the 23.6 million addresses in Nonresponse Followup, administrative records indicated that 16.2 million addresses or 68.5 percent were occupied, and 7.4 million addresses or 31.5 percent were vacant.

Table 20. 2010 Census and Administrative Records Housing Unit Status by Mode

Mode	2010 Census Housing Unit Count	Occupied in Administrative Records		Vacant in Administrative Records	
	Number	Number	Percent	Number	Percent
Total	116,716,292	96,083,076	82.3	20,633,216	17.7
Nonresponse Followup	23,584,428	16,163,930	68.5	7,420,498	31.5
Mailout/Mailback	82,315,147	72,141,619	87.6	10,173,528	12.4
Other	10,816,717	7,777,527	71.9	3,039,190	28.1

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Population Count

Table 21 shows whether the population count at an address is the same, lower, or higher in administrative records relative to the 2010 Census.

Table 21. 2010 Census and Administrative Records Population Count at an Address

	Occupied Housing Units in 2010 Census and Administrative Records	Population Count Lower in Administrative Records		Population Count the Same in Administrative Records		Population Count Higher in Administrative Records	
		Number	Percent	Number	Percent	Number	Percent
Housing units	96,083,076	17,122,713	17.8	55,469,632	57.7	23,490,731	24.4

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Of the 116.7 million 2010 Census occupied units, 96.1 million were also designated as occupied in administrative records. Of these, 55.5 million or 57.7 percent of the 2010 Census and administrative records addresses had the same population count. For 17.1 million or 17.8 percent of addresses, administrative records had a lower population count relative to the 2010 Census. For 23.5 million or 24.4 percent of addresses, administrative records had a higher population count relative to the 2010 Census.

Table 22 shows the difference in the population counts when administrative records had a higher or lower population count at an address relative to the 2010 Census.

Table 22. Difference in Population Count, when Administrative Records had a Higher or Lower Population Count Relative to the 2010 Census

Difference in Population Count, when Administrative Records had a Higher or Lower Population Count Relative to 2010 Census	Total Housing Units Occupied in Both the 2010 Census and Administrative Records, where Administrative Records had Higher or Lower Population Count		Population Count Lower in Administrative Records		Population Count Higher in Administrative Records	
	Number	Percent	Number	Percent	Number	Percent
Total	40,613,444	100.0	17,122,713	100.0	23,490,731	100.0
1	25,851,974	63.7	10,947,832	63.9	14,904,142	63.4
2	8,329,611	20.5	3,475,349	20.3	4,854,262	20.7
3	3,399,243	8.4	1,568,248	9.2	1,830,995	7.8
4	1,518,342	3.7	663,782	3.9	854,560	3.6
5	694,777	1.7	266,772	1.6	428,005	1.8
6 or More	819,497	2.0	200,730	1.2	618,767	2.6

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

When administrative records had either a lower or higher number of people at an address relative to the 2010 Census, for the majority of addresses, the administrative records population count was either higher or lower by one person. Of the 17.1 million records that administrative data had a lower population count relative to the 2010 Census, 10.9 million or 63.9 percent of these records were lower by one person. Similarly, of the 23.5 million records where administrative data had a higher population count, 63.4 percent of these records were higher by one person. About 20 percent of the records were either lower or higher by two persons. About 8 percent of the records were either lower or higher by three persons, and the percentages were successively lower for four, five, and six or more persons. Future research is needed to explore the sources and reasons for the count differences.

5.4 Demographic Quality and Coverage Assessment

Since agreements with commercial data vendors prohibit direct comparisons of data across sources, commercial file names will not be used when presenting analysis comparing the commercial data sources. Instead, commercial data files will be called commercial file 1, commercial file 2, etc. in this section. Some commercial data files do not have data for Hispanic origin, race, or sex.

Quality Assessment

This section discusses the quality of demographic characteristics in the federal and commercial files, using 2010 Census unedited demographic characteristics as the gold standard for comparison purposes. For each data source in the 2010 Census Match Study, persons were matched to the 2010 Census by PIK and then responses from the 2010 Census were compared to the demographic data provided by federal agencies and commercial data vendors.

In addition, Numident and previous census records' demographic data were evaluated, specifically the Census 2000 and 2001-2009 ACS data as these are large sources of demographic data that could be used in conjunction with other administrative data to assist in census operations.²⁴ Tax files are not included in this analysis as they do not contain demographic characteristics, and other federal files only include some demographic characteristics.

Quality of Hispanic Origin Data in Administrative Records

Table 23 shows the number and percentage of persons that had the same Hispanic origin response in administrative records and the 2010 Census by administrative records source. While the terminology “response” is used in this section, the data from some sources were modeled for Hispanic origin and race and therefore were not based on a response from a resident of the household.

Table 23. Number and Percentage of Administrative Records Hispanic Origin Response Data that Matched to the 2010 Census

2010 Census and Administrative Records Hispanic Origin Response Match by Source File	Hispanic		Not Hispanic	
	Number	Percent	Number	Percent
Federal Files				
Previous Census Records	18,137,918	93.1	162,270,334	99.4
Numident	18,898,237	54.2	215,259,972	99.7
HUD CHUMS	507,655	80.0	3,987,563	98.5
HUD PIC	1,009,383	86.0	4,405,539	98.1
HUD TRACS	14,181	78.6	105,010	98.6
TANF	220,988	70.7	1,659,036	98.3
MEDB	812,807	29.4	37,825,607	99.9
Commercial Files				
Commercial File 1	8,260,777	83.5	94,604,335	98.2
Commercial File 2	11,868,492	77.3	140,335,009	98.0
Commercial File 3	9,206,375	80.2	114,014,452	98.0
Commercial File 4	4,510,662	77.1	50,604,881	97.9

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The quality of data for non-Hispanics in federal files was considerably higher compared to Hispanics. The quality range was also less variable for non-Hispanics compared to Hispanics. The quality of Hispanic origin data in federal files ranged from 29.4 percent to 93.1 percent for the Hispanic population and 98.1 percent to 99.9 percent for the non-Hispanic population.

²⁴ For the 2010 Census, previous census records (Census 2000 and 2001-2009 ACS data) were used in race and Hispanic origin item imputation processes.

Previous census data had the highest match for Hispanic response data in federal sources at 93.1 percent, followed by HUD PIC (86.0 percent) and HUD CHUMS (80.0 percent). MEDB had the lowest percentage of Hispanic response data that matched the 2010 Census at 29.4 percent.

The percentage of data for Hispanics in commercial files that matched to the 2010 Census ranged from 77.1 percent to 83.5 percent. For Hispanics, commercial file 4 had the lowest percentage that matched to the 2010 Census and commercial file 1 had the highest percentage that matched. Similar to federal files, commercial sources also had high quality response data for non-Hispanics.

Quality of Race Responses in Administrative Records

Table 24 shows the percentage of federal and commercial race response data that matched to the 2010 Census (see Appendix 2 for numbers).

Table 24. Percentage of Administrative Records Race Response Data that Matched to the 2010 Census

2010 Census and Administrative Records Race Response Match by Source File	White Alone	Black Alone	American Indian or Alaska Native Alone	Asian Alone	Native Hawaiian or Other Pacific Islander Alone	Some Other Race Alone	Two or More Races
Federal Files							
Previous Census Records	96.8	96.2	63.2	94.1	59.7	54.9	36.3
Numident	99.1	98.3	51.4	84.3	74.4	17.7	N/A
IHS	N/A	N/A	97.6	N/A	N/A	N/A	N/A
HUD CHUMS	98.0	87.4	24.6	65.0	46.9	N/A	3.6
HUD PIC	97.2	96.3	41.7	89.3	62.5	N/A	6.9
HUD TRACS	96.1	95.4	46.9	87.3	37.0	14.5	9.7
TANF	97.6	95.9	73.0	80.9	76.0	N/A	12.8
MEDB	99.0	97.9	49.1	58.0	N/A	14.1	N/A
Commercial Files							
Commercial File 1	97.9	43.6	N/A	85.2	14.3	3.0	N/A
Commercial File 2	97.8	37.3	6.4	73.6	19.2	1.1	N/A
Commercial File 3	94.9	61.1	13.2	79.7	17.0	3.4	N/A
Commercial File 4	94.7	58.2	8.6	79.8	16.6	3.4	N/A

Note: N/A in tables in this report indicates that data were not available for a demographic group.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The White alone population had the highest quality response data in both federal and commercial files relative to other race groups. In federal files, the quality of race response data ranged from 96.1 percent to 99.1 percent for the White alone population. Commercial files had a similar though slightly lower range, from 94.7 percent to 97.9 percent. Among the federal files, Numident had the highest percentage of White alone response data that matched to the 2010 Census (99.1 percent), followed by MEDB (99.0 percent), and HUD CHUMS (98.0 percent). HUD TRACS had the lowest percentage match for White alone response at 96.1 percent. For the

commercial files, commercial file 4 had the lowest percentage match (94.7 percent) while commercial file 1 had the highest match (97.9 percent) for the White alone response.

The quality of response data was lower for the Black alone population relative to the White alone population in federal data. The quality of the race response data for the Black alone population ranged from 87.4 percent to 98.3 percent. The commercial files had a considerably lower percentage of the Black alone population that matched to the 2010 Census relative to the White alone population, a range from 37.3 percent to 61.1 percent. Among federal files, similar to the White alone population, the Numident had the highest percentage match for the Black alone population at 98.3 percent. This was followed by MEDB (97.9 percent) and HUD PIC (96.3 percent). HUD CHUMS had the lowest percentage Black alone response match to the 2010 Census (87.4 percent). Commercial file 2 had the lowest percentage match for the Black alone population, and commercial file 3 had the highest percentage that matched.

The quality of federal file race response data was considerably lower for the AIAN alone population compared to the White alone and Black alone populations. The percentage of AIAN alone race responses that matched to the 2010 Census in the federal files ranged from 24.6 percent to 97.6 percent. IHS and TANF were the two federal files that had a relatively high percentage of AIAN alone responses that matched, 97.6 percent and 73.0 percent respectively, whereas 63.2 percent of the responses in previous census records matched, and 51.4 percent or fewer of the responses for the remaining federal data sources matched the 2010 Census. Similar to the Black alone population, HUD CHUMS had the lowest percentage of AIAN matches (24.6 percent). Commercial file 1 did not have any data on the AIAN population. Among the commercial files that had data on this population, the percentages of responses that matched the 2010 Census were low, 6.4 percent to 13.2 percent. Similar to the Black alone population, commercial file 2 had the lowest percentage of AIAN alone responses that matched, and commercial file 3 had the highest.

For the federal files, the Asian alone population had higher percentages of race responses that matched the 2010 Census relative to the AIAN alone population, but lower percentages compared to the White alone and Black alone populations, 58.0 percent to 94.1 percent. Previous census records had the highest percentage match (94.1 percent), followed by HUD PIC (89.3 percent) and HUD TRACS (87.3 percent). MEDB had the lowest percentage of Asian alone responses that matched at 58.0 percent. For commercial files, the Asian alone population had higher percentages that matched the 2010 Census relative to both the Black alone and AIAN alone populations, but lower matches relative to the White alone population. The percentage of commercial data responses that matched the 2010 Census for the Asian alone population ranged from 73.6 percent to 85.2 percent. Similar to the Black alone and AIAN alone populations, commercial file 2 had the lowest match for the Asian alone population. Similar to the White alone population, commercial file 1 had the highest match.

For the NHPI alone population, 59 percent or higher of the responses in four of the seven federal datasets matched the 2010 Census. TANF had the highest percentage of NHPI alone responses that matched (76.0 percent), followed by the Numident (74.4 percent), HUD PIC (62.5 percent), and previous census records (59.7 percent). HUD TRACS had the lowest percentage of responses that matched for this population at 37.0 percent. The percentage of responses that matched in the commercial files for the NHPI alone community was considerably lower than the White alone, Black alone, and Asian alone populations, but higher than the AIAN alone population. For the NHPI alone population, 14.3 percent to 19.2 percent of the responses in the commercial files matched to the 2010 Census. Commercial file 1 had the lowest match and commercial file 2 had the highest match.

Only four of the seven federal files had a race category equivalent to SOR. Of these four data sources, MEDB had the lowest percentage of SOR alone responses that matched to the 2010 Census (14.1 percent), and previous census records had the highest percentage (54.9 percent). About 14.5 percent of HUD TRACS SOR alone responses matched to the 2010 Census and the Numident matched to 17.7 percent. These match percentages were the second lowest matches across all race groups for the federal data. In the commercial files, 1.1 percent to 3.4 percent of the SOR alone responses matched to the 2010 Census. This was the lowest match percentage of all the race groups represented in the commercial files. Similar to the majority of race groups, commercial file 2 had the lowest percentage of SOR alone responses that matched to the 2010 Census. Commercial file 3 and commercial file 4 had the highest percentages that matched.

The multiracial population had the lowest percentage of responses that matched in the federal files to the 2010 Census relative to other race groups. Of the five federal files that had data on the multiracial population, previous census records had the highest percentage that matched at 36.3 percent. HUD CHUMS had the lowest percentage that matched at 3.6 percent. TANF, HUD TRACS, and HUD PIC matched the 2010 Census multiracial population at 12.8 percent, 9.7 percent, and 6.9 percent respectively. The commercial files did not classify individuals as multiracial.

Quality of Age Responses in Administrative Records

Table 25 shows the percentage of federal and commercial age response data that matched to the 2010 Census overall and by age group (see Appendix 3 for numbers). The percentage of records that matched the age data in the 2010 Census was 95.2 percent or higher for all federal source files except HUD CHUMS. MEDB had the highest percentage match on age at 98.5 percent, followed by the Numident (97.9 percent) and SSS (97.8 percent). HUD CHUMS had the lowest age response match at 24.4 percent. The match is low because the HUD CHUMS file only included persons' year of birth, while the other files provided date of birth which more accurately can be matched to the age of persons in the 2010 Census. Relative to the federal files, the commercial files had lower percentages of age responses that matched to the 2010 Census.

Commercial file 4 had the highest percentage of age responses that matched (90.5 percent) whereas commercial file 1 had the lowest percentage match (79.0 percent).

Table 25. Percentage of Administrative Records Age Response Data that Matched to the 2010 Census

2010 Census and Administrative Records Age Response Match by Source File	Age							
	Total	0-2	3-17	18-24	25-44	45-64	65-74	75 and older
Federal Files								
Previous Census Records	95.7	85.7	94.7	95.3	95.6	96.1	96.5	96.3
Numident	97.9	97.6	98.0	97.8	98.0	98.1	98.6	96.6
IHS	96.6	95.3	96.3	96.6	97.0	96.8	96.3	94.9
HUD CHUMS	24.4	N/A	N/A	19.1	24.5	25.0	25.9	26.8
HUD PIC	97.0	95.8	96.9	97.1	97.4	97.3	97.0	96.2
HUD TRACS	96.9	96.5	96.7	97.1	97.1	96.9	97.4	96.9
SSR	95.2	97.0	97.3	97.6	97.0	95.6	92.3	89.4
SSS	97.8	N/A	N/A	98.0	97.7	N/A	N/A	N/A
TANF	96.7	96.3	96.7	97.0	97.2	96.4	94.4	88.1
MEDB	98.5	97.5	96.0	97.5	98.0	98.1	98.8	98.3
Commercial Files								
Commercial 1	79.0	0.3	6.4	77.2	78.6	79.5	79.2	78.3
Commercial 2	88.9	16.8	59.3	83.3	87.6	90.1	91.1	89.8
Commercial 3	89.1	N/A	N/A	81.4	88.6	90.2	90.8	89.9
Commercial 4	90.5	N/A	N/A	85.4	90.3	91.5	92.2	91.0
Commercial 5	88.6	N/A	3.1	79.1	90.3	88.4	87.5	86.1
Commercial 6	80.9	71.4	0.3	75.4	82.0	80.9	84.2	84.0
Commercial 7	87.2	N/A	N/A	78.6	88.1	87.5	87.1	85.9
Commercial 8	90.4	20.0	73.5	92.0	90.6	90.3	90.7	90.4

Note: N/A indicates that data were not available for a demographic group.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Across the federal files, the quality of age response data showed some slight variation according to age group. Where differences existed was in the presence of age response data. The HUD CHUMS file did not include anyone under the age of 18, MEDB included relatively few persons under 18, and the SSS file only included data on those between the ages of 18 to 25. Similarly, the commercial records had relatively few persons under the age of 18 and had lower match rates for those who were included. Commercial file 2 included more people under age 18 but the quality of the age data was lower relative to other age groups.

Quality of Sex Responses in Administrative Records

Table 26 shows the number and percentage of federal and commercial sex response data that matched to the 2010 Census.

Table 26. Number and Percentage of Administrative Records Sex Response Data that Matched to the 2010 Census

2010 Census and Administrative Records Sex Response Match by Source File	Female		Male	
	Number	Percent	Number	Percent
Federal Files				
Previous Census Records	99,582,513	99.5	91,377,033	99.5
Numident	132,710,367	99.4	125,356,726	99.4
HUD CHUMS	2,310,839	98.1	2,437,053	98.7
HUD PIC	3,742,607	99.0	2,199,033	97.9
HUD TRACS	1,341,994	98.9	695,095	98.2
IHS	1,117,176	99.4	995,603	99.2
MEDB	23,691,186	99.6	19,068,303	99.7
SSS	N/A	N/A	11,994,797	100.0
TANF	1,298,748	99.1	768,904	98.1
Commercial Files				
Commercial File 1	63,605,178	98.8	55,906,815	97.0
Commercial File 2	81,263,837	98.6	72,248,387	98.6
Commercial File 3	67,065,482	97.0	59,346,369	97.7
Commercial File 4	31,208,537	97.0	25,412,404	97.4
Commercial File 5	76,293,351	97.2	68,872,874	97.9
Commercial File 6	412,740	95.6	324,334	94.7
Commercial File 7	50,177,060	97.2	45,989,876	98.4

Note: N/A indicates that data were not available for a demographic group.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The quality of sex response data ranged from 94.7 percent to 100.0 percent across both federal and commercial files for both sexes. Among the federal administrative files, HUD PIC had the lowest percentage that matched for males at 97.9 percent. SSS had the highest match rate for males at 100.0 percent. HUD CHUMS had the lowest percentage that matched for females at 98.1 percent, and MEDB had the highest at 99.6 percent. For the commercial administrative files, commercial file 6 had the lowest percentage match for both males at 94.7 percent and females at 95.6 percent. Commercial file 2 had the highest match for males at 98.6 percent, and commercial file 1 had the highest match for females at 98.8 percent.

Demographic Coverage Assessment

This section discusses demographic characteristic coverage of the 2010 Census by the federal and commercial files, including the Numident and previous census records. Persons in the 2010 Census were matched by PIK to each data source to determine if the federal or commercial files provided any demographic data for that person on Hispanic origin, race, age, and sex regardless of the quality. This assessment indicates whether data are present for demographic groups in the 2010 Census, not whether the demographic data are the same in the 2010 Census and administrative records.

Table 27 shows whether demographic data were present in administrative records by demographic group. Administrative records had Hispanic origin response data for 278.0 million persons in the 2010 Census (90.1 percent). A higher percentage of non-Hispanics had Hispanic origin response data in administrative records relative to Hispanics. Of the 258.3 million non-Hispanics in the 2010 Census, administrative records had Hispanic origin response data for 238.2 million or 92.2 percent. Of the 50.5 million Hispanics in the 2010 Census, 39.8 million or 78.9 percent had Hispanic origin response data in administrative records.

Table 27. Coverage of 2010 Census Demographic Data by Administrative Records Demographic Response Data

Demographic Characteristics	2010 Census		Coverage of 2010 Census Demographic Data by Administrative Records Demographic Response Data	
	Number	Percent	Number	Percent
Total Population	308,745,538	100.0	278,484,228	90.2
Hispanic or Latino Origin	308,745,538	100.0	278,045,021	90.1
Hispanic	50,477,594	100.0	39,814,879	78.9
Not Hispanic	258,267,944	100.0	238,230,142	92.2
Race	308,745,538	100.0	239,489,480	77.6
White Alone	223,553,254	100.0	181,023,292	81.0
Black Alone	38,929,315	100.0	30,456,240	78.2
American Indian or Alaska Native Alone	2,932,370	100.0	2,256,067	76.9
Asian Alone	14,674,336	100.0	10,839,299	73.9
Native Hawaiian or Other Pacific Islander Alone	540,064	100.0	377,663	69.9
Some Other Race Alone	19,107,368	100.0	8,799,778	46.1
Two or More Races	9,008,831	100.0	5,737,141	63.7
Age	308,745,538	100.0	278,123,833	90.1
0-2	12,019,146	100.0	10,771,945	89.6
3-17	62,162,321	100.0	56,522,460	90.9
18-24	30,646,519	100.0	26,032,464	84.9
25-44	82,123,330	100.0	71,307,164	86.8
45-64	81,499,596	100.0	75,632,822	92.8
65-74	21,727,578	100.0	20,498,121	94.3
75 and older	18,567,048	100.0	17,358,857	93.5
Sex	308,745,538	100.0	278,038,511	90.1
Male	151,775,099	100.0	135,515,017	89.3
Female	156,970,439	100.0	142,523,494	90.8

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Data on race were available for 239.5 million or 77.6 percent of 2010 Census respondents. Administrative records provided the greatest level of race data coverage for the White alone population and the lowest level of race data coverage for the SOR alone population in the 2010

Census. Of the 223.6 million persons classified as White alone in the 2010 Census, 181 million or 81.0 percent had race data in administrative records. The next highest level of race data coverage was for the Black alone population (78.2 percent), followed by AIAN alone (76.9 percent), Asian alone (73.9 percent), NHPI alone (69.9 percent), and the Two or More Races population (63.7 percent). Administrative records contained race data for just under half (46.1 percent) of the SOR alone population.

Administrative records provided coverage of age data for 278.1 million or 90.1 percent of all persons on the 2010 Census. Coverage by age group in the 2010 Census ranged from 84.9 percent to 94.3 percent with older age groups more likely to have age data present in administrative records relative to younger age groups. Age coverage by administrative records was greatest for those in the 65 to 74 age group (94.3 percent), followed by 75 and older (93.5 percent), 45 to 64 (92.8 percent), 3 to 17 (90.9), 0 to 2 (89.6 percent), and 25 to 44 (86.8 percent) age groups. The age group with the lowest coverage was those aged 18 to 24 at 84.9 percent.

Sex data were available in the administrative records for 278.0 million or 90.1 percent of all persons on the 2010 Census. Coverage was slightly higher for females in the 2010 Census than for males. For females in the 2010 Census, 90.8 percent had data on sex in administrative data. For males in the 2010 Census, 89.3 percent also had data on sex in administrative records.

Coverage by Mode by Demographic Group

Table 28 shows whether Hispanic origin data were present in administrative records by mode.

Table 28. Coverage of 2010 Hispanic Origin Data by Administrative Records Hispanic Origin Response Data by Mode

Coverage of 2010 Hispanic Origin Data by Administrative Records Hispanic Origin Response Data	2010 Census		Administrative Records	
	Number	Percent	Number	Percent
NRFU	60,432,209	100.0	48,868,213	80.9
Hispanic	12,474,326	100.0	8,673,291	69.5
Not Hispanic	47,957,883	100.0	40,194,922	83.8
Mailout/Mailback	205,816,623	100.0	198,842,905	96.6
Hispanic	28,619,508	100.0	25,137,353	87.8
Not Hispanic	177,197,115	100.0	173,705,552	98.0
Other	42,496,706	100.0	30,333,903	71.4
Hispanic	9,383,760	100.0	6,004,235	64.0
Not Hispanic	33,112,946	100.0	24,329,668	73.5

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Administrative data had Hispanic origin response data for 96.6 percent of persons whose response was obtained via Mailout/Mailback in the 2010 Census. There were 80.9 percent of NRFU respondents in the 2010 Census that had Hispanic origin response data in administrative records. Administrative records covered about 10 percent more of the non-Hispanic population compared to the Hispanic population regardless of mode.

Table 29 shows whether race data were present in administrative records by mode.

Table 29. Coverage of 2010 Race Data by Administrative Records Race Response Data by Mode

Coverage of 2010 Race Data by Administrative Records Race Response Data				
	2010 Census		Administrative Records	
	Number	Percent	Number	Percent
NRFU	60,432,209	100.0	40,596,485	67.2
White Alone	38,193,839	100.0	27,279,990	71.4
Black Alone	9,665,248	100.0	6,828,025	70.6
American Indian or Alaska Native Alone	614,416	100.0	466,097	75.9
Asian Alone	2,935,599	100.0	1,836,450	62.6
Native Hawaiian or Other Pacific Islander Alone	170,657	100.0	111,437	65.3
Some Other Race Alone	6,574,514	100.0	2,658,243	40.4
Two or More Races	2,277,936	100.0	1,416,243	62.2
Mailout/Mailback	205,816,623	100.0	173,992,345	84.5
White Alone	158,738,870	100.0	137,265,519	86.5
Black Alone	22,179,559	100.0	19,345,638	87.2
American Indian or Alaska Native Alone	1,367,303	100.0	1,117,407	81.7
Asian Alone	9,415,785	100.0	7,614,070	80.9
Native Hawaiian or Other Pacific Islander Alone	234,376	100.0	190,972	81.5
Some Other Race Alone	8,587,123	100.0	4,868,058	56.7
Two or More Races	5,293,607	100.0	3,590,681	67.8
Other	42,496,706	100.0	24,900,650	58.6
White Alone	26,620,545	100.0	16,477,783	61.9
Black Alone	7,084,508	100.0	4,282,577	60.4
American Indian or Alaska Native Alone	950,651	100.0	672,563	70.7
Asian Alone	2,322,952	100.0	1,388,779	59.8
Native Hawaiian or Other Pacific Islander Alone	135,031	100.0	75,254	55.7
Some Other Race Alone	3,945,731	100.0	1,273,477	32.3
Two or More Races	1,437,288	100.0	730,217	50.8

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Data on race collected via Mailout/Mailback was most likely to be covered by administrative records (84.5 percent), followed by NRFU (67.2 percent), and other response operations (58.6 percent).

All race categories had the highest levels of coverage in Mailout/Mailback mode. Administrative records race response coverage for 2010 Census respondents in Mailout/Mailback ranged from a high of 87.2 percent for the Black alone population to a low of 56.7 percent for SOR alone. Persons reporting White alone via Mailout/Mailback had the second highest administrative records race coverage (86.5 percent), followed by AIAN alone (81.7 percent), NHPI alone (81.5 percent), Asian alone (80.9), and Two or More Races (67.8 percent).

The coverage rate for race responses collected via NRFU was highest for AIAN alone. Of the approximately 614,000 persons who reported AIAN alone in NRFU, administrative records had race data for approximately 466,000 or 75.9 percent of respondents. The White alone population had the next highest coverage rate (71.4 percent), followed by the Black alone (70.6 percent), NHPI alone (65.3 percent), Asian alone (62.6 percent), and Two or More Races (62.2 percent) populations in NRFU. The SOR alone population had the lowest coverage rate in NRFU at 40.4 percent.

Table 30 shows whether age data were present in administrative records by mode.

Of 2010 Census respondents with a PIK in Mailout/Mailback, 96.6 percent had age data in administrative records. Age response coverage was lower for NRFU (80.9 percent) and other modes (71.4 percent). Among NRFU respondents, the age groups 3 to 17 and 0 to 2 had the highest age response coverage in administrative records at 84.8 percent and 83.3 percent respectively. This was followed by age groups 45 to 64 (82.3 percent), 65 to 74 (81.5 percent), and 75 and older (80.3 percent). Administrative record coverage for age data was lowest in NRFU for the 18 to 24 (76.4 percent) and 25 to 44 (78.4 percent) age groups.

Table 30. Coverage of 2010 Age Data by Administrative Records Age Response Data by Mode

Coverage of 2010 Age Data by Administrative Records Age Response Data	2010 Census		Administrative Records	
	Number	Percent	Number	Percent
NRFU	60,432,209	100.0	48,867,705	80.9
0-2	2,713,417	100.0	2,261,220	83.3
3-17	13,959,494	100.0	11,839,104	84.8
18-24	7,322,346	100.0	5,594,530	76.4
25-44	19,498,293	100.0	15,293,653	78.4
45-64	12,498,785	100.0	10,283,837	82.3
65-74	2,447,491	100.0	1,994,655	81.5
75 and older	1,992,383	100.0	1,600,706	80.3
Mailout/Mailback	205,816,623	100.0	198,892,28	96.6
0-2	7,161,233	100.0	6,894,732	96.3
3-17	39,058,528	100.0	37,769,055	96.7
18-24	15,982,399	100.0	15,164,326	94.9
25-44	52,033,098	100.0	49,279,405	94.7
45-64	60,075,645	100.0	58,829,943	97.9
65-74	17,372,352	100.0	17,101,837	98.4
75 and older	14,133,368	100.0	13,852,987	98.0
Other	42,496,706	100.0	30,363,843	71.4
0-2	2,144,496	100.0	1,615,993	75.4
3-17	9,144,299	100.0	6,914,301	75.6
18-24	7,341,774	100.0	5,273,608	71.8
25-44	10,591,939	100.0	6,734,106	63.6
45-64	8,925,166	100.0	6,519,042	73.0
65-74	1,907,735	100.0	1,401,629	73.5
75 and older	2,441,297	100.0	1,905,164	78.0

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Table 31 shows whether sex data were present in administrative records by mode. Similar to other demographic characteristics, administrative record coverage was highest for sex in the Mailout/Mailback universe (96.6 percent) and lower in the NRFU universe (80.9 percent) and via other modes (71.4 percent). For each of the three response mode categories, administrative record coverage of females in the 2010 Census was slightly higher than for males.

**Table 31. Coverage of 2010 Sex Data by Administrative Records
Sex Response Data by Mode**

Coverage of 2010 Sex Data by Administrative Records Sex Response Data	2010 Census		Administrative Records	
	Number	Percent	Number	Percent
NRFU	60,432,209	100.0	48,866,122	80.9
Male	30,490,505	100.0	24,352,259	79.9
Female	29,941,704	100.0	24,513,863	81.9
Mailout/Mailback	205,816,623	100.0	198,838,820	96.6
Male	99,125,339	100.0	95,563,261	96.4
Female	106,691,284	100.0	103,275,559	96.8
Other	42,496,706	100.0	30,333,569	71.4
Male	22,159,255	100.0	15,599,497	70.4
Female	20,337,451	100.0	14,734,072	72.4

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Coverage by Source File

Table 32 shows administrative records coverage of Hispanic origin response data by federal and commercial data source. For all demographic characteristics, the size of the source file strongly influenced the coverage of demographic data in the 2010 Census.

The range of coverage for the 2010 Census Hispanic population in federal data was 0.04 percent to 78.1 percent and 13.0 percent to 33.9 percent for commercial data. The Numident file had the highest percent coverage of Hispanic origin response data for the Hispanic population (78.1 percent) and non-Hispanic population (92.1 percent). Previous census records had the second highest coverage at 43.1 percent for the Hispanic population and 69.1 percent for non-Hispanics. HUD TRACS had the lowest coverage for Hispanics (0.04 percent) and non-Hispanics (0.05 percent).

Commercial data sources covered 13.0 percent to 33.9 percent of Hispanics in the 2010 Census and 22.0 percent to 60.6 percent of non-Hispanics. Among the commercial sources, commercial file 2 provided the highest level of Hispanic origin response coverage at 33.9 percent for Hispanics and 60.6 percent for non-Hispanics.

Table 32. Number and Percent Coverage of 2010 Hispanic Origin Data by Administrative Records Source Files

Coverage of 2010 Hispanic Origin Data by Administrative Records Hispanic Origin Response Data by Source	Hispanic		Not Hispanic	
	Number	Percent	Number	Percent
2010 Census	50,477,594	100	258,267,944	100
Federal Files				
Previous Census Records	21,764,183	43.1	178,348,197	69.1
Numident	39,399,214	78.1	237,807,990	92.1
HUD CHUMS	697,169	1.4	4,316,851	1.7
HUD PIC	1,364,197	2.7	5,377,679	2.1
HUD TRACS	20,987	0.0	127,991	0.0
MEDB	3,070,925	6.1	42,825,729	16.6
TANF	365,626	0.7	1,963,550	0.8
Commercial Files				
Commercial File 1	11,349,460	22.5	122,397,813	47.4
Commercial File 2	17,093,059	33.9	156,451,837	60.6
Commercial File 3	12,732,083	25.2	126,544,309	49.0
Commercial File 4	6,540,972	13.0	56,774,215	22.0

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Table 33 shows administrative records coverage of race response data by federal and commercial data source (see Appendix 4 for numbers).

With few exceptions, the White alone population had higher coverage rates across the sources relative to other race groups. The coverage rate for the White alone population ranged from 0.1 percent to 71.1 percent in federal data and 18.5 percent to 55.2 percent in commercial data. Previous census records had the highest coverage rate (71.1 percent) for the White alone population, followed by the Numident at 66.0 percent. The IHS had the lowest coverage rate at 0.1 percent. Of commercial files, commercial file 4 had the lowest coverage rate for the White alone population, and commercial file 2 had the highest. Commercial file 4 had the lowest coverage rate and commercial file 2 had the highest for all race groups.

The coverage rate for the Black alone population ranged from 0.03 percent to 66.9 percent across federal sources and 16.9 percent to 45.9 percent in commercial data sources. The Numident had the highest coverage rate for the Black alone population (66.9 percent), followed by previous census records (57.0 percent). Similar to the White alone population, IHS had the lowest coverage rate at 0.03 percent.

Table 33. Percent Coverage of 2010 Race Data by Administrative Records Source Files

Coverage of 2010 Race Data by Administrative Records Race Response Data by Source	White Alone	Black Alone	American Indian or Alaska Native Alone	Asian Alone	Native Hawaiian or Other Pacific Islander Alone	Some Other Race Alone	Two or More Races
Federal Files							
Previous Census Records	71.1	57.0	55.3	49.5	43.7	32.7	48.3
Numident	66.0	66.9	54.0	64.1	57.6	6.7	36.8
HUD CHUMS	1.7	1.4	1.0	1.0	1.2	1.1	1.0
HUD PIC	1.1	8.3	2.9	1.2	3.7	2.6	4.0
HUD TRACS	0.4	2.1	0.8	0.6	0.5	0.6	1.0
IHS	0.1	0.0	41.4	0.0	0.1	0.1	2.7
MEDB	16.9	11.6	9.2	8.5	6.4	2.4	5.3
TANF	0.5	2.0	2.3	0.2	3.2	0.6	1.7
Commercial Files							
Commercial File 1	38.9	28.9	15.5	27.0	12.1	2.1	14.4
Commercial File 2	55.2	45.9	35.8	41.6	27.0	27.8	31.1
Commercial File 3	43.6	32.6	22.6	29.9	16.5	2.6	18.2
Commercial File 4	18.5	16.9	10.6	13.0	8.2	1.3	9.2

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

For the AIAN alone population, the federal coverage rate ranged from 0.8 percent to 55.3 percent, and the commercial coverage rate ranged from 10.6 percent to 35.8 percent. Similar to the White alone and Black alone populations, the Numident and previous census records had the two highest coverage rates. Previous census records had the highest coverage at 55.3 percent, and the Numident had the second highest coverage rate at 54.0 percent. IHS also covered a relatively substantial proportion of response data at 41.4 percent. HUD TRACS had the lowest coverage rate at 0.8 percent.

The federal coverage rate for the Asian alone population ranged from 0.01 percent to 64.1 percent, and the commercial coverage rate ranged from 13.0 percent to 41.6 percent. As with the previous race groups discussed, the Numident and previous census records had the greatest coverage for the Asian alone population. The Numident covered 64.1 percent of the 2010 Census Asian alone population, and previous census records covered 49.5 percent.

For the NHPI population, the federal coverage rate range was 0.1 percent to 57.6 percent, and the commercial coverage rate range was 8.2 percent to 27.0 percent. The Numident and previous census records had the highest coverage rates, 57.6 percent and 43.7 percent respectively. IHS had the lowest coverage rate at 0.1 percent.

With few exceptions, the SOR alone population had the lowest coverage rates across federal and commercial data sources. The federal coverage rate ranged from 0.1 percent to 32.7 percent, and

commercial data covered 1.3 percent to 27.8 percent. Previous census records had the highest coverage at 32.7 percent, followed by the Numident at 6.7 percent.

The federal coverage rate for the Two or More Races population ranged from 1.0 percent to 48.3 percent, and the commercial coverage rate ranged from 9.2 percent to 31.1 percent. Previous census records covered 48.3 percent of the 2010 Census Two or More Races population, followed by the Numident at 36.8 percent.

Table 34 shows administrative records coverage of age response data by federal and commercial data source (see Appendix 5 for numbers).

Table 34. Percent Coverage of 2010 Age Data by Administrative Records Source Files

Coverage of 2010 Age Data by Administrative Records Age Response Data by Source	Age						
	0-2	3-17	18-24	25-44	45-64	65-74	75 and older
Federal Files							
Previous Census Records	1.5	42.1	69.9	67.7	79.9	84.0	85.0
Numident	89.5	90.9	84.9	86.8	92.8	94.3	93.5
HUD CHUMS	0.0	0.0	0.9	3.5	1.7	1.0	0.5
HUD PIC	2.9	4.1	2.3	1.7	1.4	1.5	1.6
HUD TRACS	1.2	0.9	0.7	0.4	0.5	1.3	2.2
IHS	0.6	0.9	0.9	0.8	0.7	0.6	0.4
MEDB	0.0	0.0	0.5	2.2	8.9	90.5	91.4
SSR	0.8	1.7	1.9	1.7	2.8	3.9	4.4
SSS	0.0	0.1	37.4	2.1	0.0	0.0	0.0
TANF	2.9	1.7	0.9	0.6	0.2	0.1	0.0
Commercial Files							
Commercial File 1	0.1	0.1	9.7	31.7	49.8	55.7	53.4
Commercial File 2	0.2	0.9	25.9	39.2	65.1	70.7	66.9
Commercial File 3	0.1	0.5	25.9	42.2	62.2	67.9	65.3
Commercial File 4	0.0	0.2	12.7	22.9	25.1	26.5	26.0
Commercial File 5	0.1	0.1	1.7	45.4	74.1	77.2	73.3
Commercial File 6	0.0	0.0	0.2	0.6	0.4	0.1	0.0
Commercial File 7	0.0	0.1	4.5	24.4	48.9	58.9	59.6
Commercial File 8	0.1	0.1	1.2	40.0	66.4	67.8	62.2

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The Numident age response data covered a substantial proportion of 0 to 2 year olds in the 2010 Census at 89.5 percent. All other federal and commercial data sources covered less than 3 percent each. The Numident age response data also covered a substantial proportion of those 3 to 17 years old in the 2010 Census at 90.9 percent. Previous census records covered 42.1 percent. All other data covered less than 4.2 percent.

The Numident covered 84.9 percent of the 18 to 24 age group, followed by previous census records at 69.9 percent. Commercial file 2 and commercial file 3 had the highest coverage rates

for the 18 to 24 age group among commercial files at 25.9 percent each. The Numident covered 86.8 percent of the 25 to 44 age group, and previous census records covered 67.7 percent. Commercial file 5 and commercial file 3 had the highest coverage for this age group at 45.4 percent and 42.2 percent respectively. The Numident covered the highest percentage of age responses for the age groups 45 to 64, 65 to 74, and 75 and older. Commercial file 5 had the highest coverage for all three of these age groups among the commercial data.

Table 35 shows administrative records coverage of sex response data by federal and commercial data source.

Table 35. Number and Percent Coverage of 2010 Sex Data by Administrative Records Source Files

Coverage of 2010 Sex Data by Administrative Records Sex Response Data by Source	Male		Female	
	Number	Percent	Number	Percent
2010 Census	151,775,099	100.0	156,970,439	100.0
Federal Files				
Previous Census Records	98,056,250	64.6	106,215,624	67.7
Numident	135,270,982	89.1	142,309,140	90.7
HUD CHUMS	2,607,763	1.7	2,472,911	1.6
HUD PIC	2,535,359	1.7	4,212,021	2.7
HUD TRACS	817,071	0.5	1,537,300	1.0
IHS	1,096,844	0.7	1,219,319	0.8
MEDB	20,481,253	13.5	25,415,378	16.2
SSS	13,211,190	8.7	70,171	0.0
TANF	874,359	0.6	1,455,174	0.9
Commercial Files				
Commercial File 1	60,929,919	40.1	67,686,745	43.1
Commercial File 2	77,959,856	51.4	87,230,172	55.6
Commercial File 3	64,421,794	42.4	73,002,480	46.5
Commercial File 4	27,970,377	18.4	34,345,858	21.9
Commercial File 5	74,729,309	49.2	82,823,112	52.8
Commercial File 6	370,287	0.2	462,087	0.3
Commercial File 7	49,329,354	32.5	54,231,282	34.5

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

The federal coverage rate for males ranged from 0.5 percent to 89.1 percent, and the commercial coverage rate ranged from 0.2 percent to 51.4 percent. For females, the federal data covered 0.04 percent to 90.7 percent, and commercial data covered 0.3 percent to 55.6 percent. Numident had the highest coverage for both males (89.1 percent) and females (90.7 percent). Previous census records had the second highest coverage at 64.6 percent for males and 67.7 percent for females. Commercial file 6 had the lowest coverage rate for males and females, and commercial file 2 had the highest coverage.

Overall, the Numident and previous census records had the highest coverage of demographic response data across all demographic groups. For many demographic groups such as age and sex, and for the White alone, Black alone, and Asian alone populations, these datasets also tended to have relatively high quality response data.

6. Related Census Program for Evaluations and Experiments Reports

The following Census Program for Evaluations and Experiments reports are related to the Census Match Study.

- 2010 Census Evaluation of Small Multi-Unit Structures Report
- 2010 Census Nonresponse Followup Operations Assessment
- 2010 Census Operational Assessment for Type of Enumeration Area Delineation

7. Lessons Learned, Conclusions, and Research Implications

Administrative records data are **available** for use in census operations. Data with a reference date appropriate for Census Day can be identified. Once acquired, federal and commercial data can be processed, unduplicated, and ready to use for Census purposes within two to four weeks. The 2010 Census Match Study used twenty files from eight federal agencies and five commercial data vendors.

Administrative records data are **reliable** for address and count confirmation for persons and addresses. Administrative records data confirmed person data in the 2010 Census for 273.6 million persons or 98.0 percent of census respondents with a PIK. Administrative records failed to match 2010 Census records lacking name data. 2010 Census address data were confirmed for 122.0 million addresses or 92.6 percent.

Administrative records data can **improve** census data quality for respondent characteristics and treatments of missing data. Census data quality can be improved by integrating administrative records information into item imputation methods. For instance, administrative data, including the Numident and previous census records had high quality age and sex data. They also contain age and sex response data for about 278 million persons in the 2010 Census.

Administrative data use can **reduce the cost** of future data collections by enhancing the MAF and through strategies addressing non-response. Administrative records can help direct field operations to areas with low person and address confirmation and areas of new construction with an operation such as targeted address canvassing. Administrative records data can be used to confirm housing unit status and to identify or confirm occupied status. Of the 116.7 million occupied housing units in the 2010 Census, administrative records showed agreement for 96.1 million. Administrative records show potential for use in household size imputations. About 55.5 million housing units in the 2010 Census had the same population count in administrative

records. When administrative records and the 2010 Census differed in population count for housing units, the population count differed by one person for 63.7 percent of those housing units.

The 2010 Census Match Study was designed to compile an unduplicated list of administrative records addresses and persons. The lists were counted and compared to 2010 Census results. Validated records in the lists were matched by unique housing unit and person identifiers. This matching was done to assess the quality and coverage of information in the administrative records files, but in essence simulated an administrative records census. The results indicate that the United States does not have an administrative records infrastructure suitable for an administrative records census. The 2010 Census Match Study sought to find the same person in the same housing unit, but the complexities of the administrative records data made the comparisons difficult and at times suspect. The complexities ranged from missing and false name, address, and date of birth information to data universe and timing discrepancies. Efforts to mitigate these challenges, such as only including validated (PIKed) persons in the comparisons introduced new problems, since the PIKed persons are likely different from the unPIKed persons in terms of characteristics and response propensities. Yet overall, the results indicate sufficient promise in administrative records to pursue operational designs for future frames and censuses.

The 2010 Census Match Study should be viewed as a national-level proof of concept for household administrative records, demonstrating the Census Bureau's ability to acquire and process public and private administrative records. The results indicate that the additions to the federal files used in StARS were worthwhile. Commercial data were a useful addition for address coverage, but more work is needed to understand how the data can enhance person coverage or person follow up operations. The unduplicated administrative records files provide high coverage, high quality information to inform occupied status imputations, and more work is needed to explore how administrative records data can be used in household size imputations without creating overcounts and undercounts of key populations. The demographic data quality analyses revealed that administrative records files contain high coverage, high quality information on age and sex, and that federal files can enhance previously collected census data for race and Hispanic origin information.

Future operational uses of administrative data need to focus on which files are fit for their particular needs; this study's results indicate that neither one file nor one composite will be adequate for both item and count imputations.

Research Implications

1. **Administrative records can enhance, but not replace the decennial census.** While the quality and coverage of administrative records relative to the 2010 Census suggests that

administrative records can be utilized in decennial census operations, the quality is not high enough and the coverage is not expansive enough to replace a traditional census.

2. Use of administrative records in Nonresponse Followup can reduce costs.

Administrative records cover a substantial number of Nonresponse Followup addresses and persons, and nearly half of person-address pairs. Of the 23.6 million addresses that responded in Nonresponse Followup in the 2010 Census, administrative records matched to 21.0 million or 89.2 percent. Administrative records also matched to a substantial number of persons that were in Nonresponse Followup in the 2010 Census. Of the 60.4 million persons in Nonresponse Followup in the 2010 Census, 48.0 million or 79.5 percent were in administrative records. Administrative records matched to a lower number and proportion of person-address pairs in Nonresponse Followup compared to addresses and persons. Of the 60.4 million 2010 person-address pairs in Nonresponse Followup, there were 28.7 million or 47.5 percent that matched to administrative records. Research and improvements in record linkage, refinements of the best address model, and acquiring data that cover those most likely to be in Nonresponse Followup may enhance the person-address match between the 2010 Census and administrative records.

3. Administrative records can assist in determining housing unit and occupancy status.

Administrative records can assist to verify whether a housing unit is a valid livable housing unit and whether it is occupied. Occupancy status results demonstrate the value of administrative records for these purposes. Of the 116.7 million occupied housing units in the 2010 Census, administrative records indicated that 96.1 million or 82.3 percent were occupied. The 2010 Census designated 15.0 million housing units as vacant, of which administrative records found that 11.4 million or 76.1 percent were not occupied. Of the 4.9 million housing units designated as deletes in the 2010 Census, administrative records indicated that 4.2 million or 85.4 percent were not occupied.

4. Administrative records can inform household population count assignment.

Administrative records had the same population count for the majority of 2010 Census housing units that matched to administrative records. Of the 116.7 million 2010 Census occupied housing units, 96.1 million matched to administrative records. Of these, 55.5 million or 57.7 percent of housing units had the same population count. When administrative records and the 2010 Census did not have the same population count, the count differed by one person for 63.7 percent of the housing units. Further research should be conducted on this universe.

5. Acquiring additional federal, state, and commercial data can improve address, person, and demographic characteristic coverage. Administrative data do not cover children as well as they cover adults. Also, the quality of race and Hispanic origin response data from federal and commercial sources varies considerably by race and Hispanic origin group. The Census Bureau should partner with federal agencies, state

agencies, community groups, and other organizations to obtain data that contain information on children living in households, and additional race and Hispanic origin response data should be acquired, particularly for groups where the quality of race or Hispanic origin response data is low in administrative records. Obtaining data for the following groups should be a priority: Two or More Races, Native Hawaiian or Other Pacific Islander, and American Indian or Alaska Native.

6. **Administrative records can inform race and Hispanic origin determination.** For some race and Hispanic origin groups, the quality of administrative records response data was high. For instance, the White alone, Black alone, and Asian alone populations had relatively high quality race response data in administrative records compared to other race groups. The quality of administrative records files ranged from 94.7 percent to 99.1 percent for the White alone population. The quality of federal data for the Black alone population ranged from 87.4 percent to 98.3 percent. The range was considerably lower for commercial data. For the Asian alone population, the quality of both federal and commercial data ranged from 58.0 percent to 94.1 percent. Data could also be used for other race groups from administrative records, but the quality was generally lower. Research should be conducted on how administrative records can assist with race and Hispanic origin determination for censuses and surveys.
7. **Administrative records can assist age and sex determination.** The quality of age and sex response data in administrative records is high. For sex, the quality of administrative data ranged from 94.7 percent to 100.0 percent across administrative records files. For age, in data sources that contained date of birth, the quality of administrative records ranged from 79.0 percent to 98.5 percent. Research should be conducted on how administrative data can assist with age and sex determination for censuses and surveys.
8. **Conduct additional record linkage research with the aim of improving match results for unvalidated person records.** Many improvements were made to the Person Identification Validation System to enhance the assignment of protected identification keys and master address file identification numbers to administrative records data. Continued record linkage research on the Person Identification Validation System should be conducted to further enhance the assignment of protected identification keys and master address file identification numbers to persons and addresses, potentially increasing the universe of persons and addresses that can be matched and unduplicated between censuses and surveys and administrative records. For instance, of the 308.7 million persons in the 2010 Census, 29.6 million did not receive a protected identification key. Of these, 10.3 million could not be sent through Person Identification Validation System processing because they lacked name and date of birth, and 19.3 million went through Person Identification Validation System processing but failed to receive a protected identification key. Additional research should be conducted on how to minimize this latter universe.

9. **Conduct record linkage research to improve match results for records with incomplete name and date of birth data.** Commercial data sources often lack complete name and date of birth information. Research to unduplicate these records that failed the Person Identification Validation System, and assess the quality of the data is needed. Research on how to use records that lack personally identifiable information is needed, moving the matching approach beyond validation using the Social Security Administration Numerical Identification File.
10. **Conduct record linkage research that improves person record unduplication.** Current record linkage techniques must determine whether two people that look similar are indeed the same person or if they are two different people. Refinements on record linkage techniques will help to more accurately unduplicate person records.
11. **Develop partnerships with federal and state agencies to better understand administrative records and enhance record linkage research.** Partnering with federal and state agencies will facilitate knowledge sharing on the availability of data that could enhance record linkage processes. This knowledge sharing will also benefit administrative records research. For instance, a better understanding of how data were collected could assist in the validation and unduplication process and improve understanding of resulting linkages.
12. **Assess whether an administrative records composite improves missing data assignment.** Building an administrative records composite involves unduplicating records, assigning persons at multiple addresses to one address, and assigning one characteristic to people that have different characteristics across source files. Research should assess the quality of missing data assignment using a composite compared to using all available administrative data.
13. **Analyze linked survey data, especially the American Community Survey, to explore characteristics associated with data coverage and consistency.** Evaluating administrative records relative to the 2010 Census provided important information, at different levels of geography and by certain characteristics, about the quality and coverage of administrative data. Other evaluations using survey data such as the American Community Survey can provide additional insights because the American Community Survey has many additional characteristics that can be analyzed.

8. Acknowledgements

The authors of this report would like to thank Deborah Wagner, Damon Smith, Matthew Bouch, Mary Layne, Juan Carlos Humud, and Michael Moldoff for PVSing the files used in the 2010 Census Match Study.

We would also like to thank Brian Clark for assisting with quality controls for the report.

9. References

- Alvey, Wendy and Fritz Scheuren. 1982. "Background for an Administrative Record Census," Proceedings of the Social Statistics Section, American Statistical Association.
- Berning, Michael. 2003. "Administrative Records Experiment in 2000 (AREX 2000)," U.S. Census Bureau.
- Bye, Barry and Dean Judson. 2004. "Results From the Administrative Records Experiment in 2000," Census 2000 Synthesis Report No. 16, U.S. Census Bureau.
- Edmonston, Barry and Charles Schultze, eds. 1995. *Modernizing the U.S. Census*, National Academy Press, Washington, D.C.
- Farber, James and Charlene Leggieri. 2002. "Building and Validating a National Administrative Records Database for the United States," New Zealand Conference on Database Integration.
- Humes, Karen, Nicholas Jones, and Roberto Ramirez. 2011. "Overview of Race and Hispanic Origin: 2010, U.S. Census Bureau," 2010 Census Briefs, C2010BR-02. Downloaded from (www.census.gov/prod/cen2010/briefs/c2010br-02.pdf).
- Johanson, Carrie, Mark Scheu, and Keith Wechter. 2011. "2010 Census Operational Assessment for Type of Enumeration Area Delineation," 2010 Census Program for Evaluations and Experiments, 2010 Census Planning Memoranda Series No. 164, U.S. Census Bureau.
- Judson, Dean and Barry Bye. 2003. "Synthesis of Results from the Administrative Records Experiment in 2000 (AREX 2000)," U.S. Census Bureau.
- Mule, Thomas. 2012. "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01, U.S. Census Bureau.
- Ralphs, Martin and Paul Tutton. 2011. "Beyond 2011: International Models for Census Taking: Current Processes and Future Developments," Beyond 2011 Project, Office for National Statistics. Version 1.0.
- Reamer, Andrew. 2010. "Counting for Dollars: The Role of the Decennial Census in the Geographic Distribution of Federal Funds," Brookings. Downloaded from (<http://www.brookings.edu>).
- Schellhamer, Teresa. 2012. "2010 Census Match Study: Commercial Data Analysis and Evaluation," Center for Administrative Records Research and Applications Internal Memorandum, U.S. Census Bureau.
- Steffey, Duane L., and Norman M. Bradburn, eds. 1994. *Counting People in the Information Age*, National Academy Press, Washington, D.C.

Virgile, Matthew. 2012. "2010 Census Evaluation of Small Multi-Unit Structures Report," 2010 Census Program for Evaluations and Experiments, 2010 Census Planning Memoranda Series No. 175, U.S. Census Bureau.

Wagner, Deborah and Mary Layne. 2012. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research & Applications' (CARRA) Record Linkage Software," Center for Administrative Records Research and Applications Internal Paper, U.S. Census Bureau.

Walker, Shelly, Susanna Winder, Geoff Jackson, and Sarah Heibel. 2012. "2010 Census Nonresponse Followup Operations Assessment," 2010 Census Program for Evaluations and Experiments, 2010 Census Planning Memoranda Series No. 190, U.S. Census Bureau.

Appendix 1. 2010 Census and Administrative Records Address Count and Match Numbers and Ratios by State

State	2010 Census Address Count	Administrative Records Address Count	2010 Census and Administrative Records Address Match	2010 Census and Administrative Records Address Count Ratio	2010 Census and Administrative Records Address Match Ratio
Total	131,704,730	151,277,043	121,967,283	114.9	92.6
Alabama	2,171,853	2,631,070	1,958,393	121.1	90.2
Alaska	306,967	284,581	216,396	92.7	70.5
Arizona	2,844,526	3,181,603	2,579,685	111.9	90.7
Arkansas	1,316,299	1,563,799	1,187,413	118.8	90.2
California	13,680,081	15,636,385	13,048,320	114.3	95.4
Colorado	2,212,898	2,548,541	2,053,765	115.2	92.8
Connecticut	1,487,891	1,657,153	1,413,863	111.4	95.0
Delaware	405,885	498,142	374,029	122.7	92.2
District of Columbia	296,719	350,341	285,015	118.1	96.1
Florida	8,989,580	10,626,269	8,517,678	118.2	94.8
Georgia	4,088,801	4,981,082	3,768,449	121.8	92.2
Hawaii	519,508	577,083	446,412	111.1	85.9
Idaho	667,796	738,029	586,574	110.5	87.8
Illinois	5,296,715	6,139,013	4,998,755	115.9	94.4
Indiana	2,795,541	3,257,283	2,675,370	116.5	95.7
Iowa	1,336,417	1,585,541	1,290,013	118.6	96.5
Kansas	1,233,215	1,451,380	1,177,071	117.7	95.4
Kentucky	1,927,164	2,272,290	1,740,059	117.9	90.3
Louisiana	1,964,981	2,364,806	1,820,451	120.3	92.6
Maine	721,830	771,113	581,006	106.8	80.5
Maryland	2,378,814	2,681,983	2,279,666	112.7	95.8
Massachusetts	2,808,254	3,164,933	2,644,298	112.7	94.2
Michigan	4,532,233	5,180,638	4,272,367	114.3	94.3
Minnesota	2,347,201	2,647,850	2,207,463	112.8	94.0
Mississippi	1,274,719	1,580,466	1,137,206	124.0	89.2
Missouri	2,712,729	3,145,375	2,492,094	115.9	91.9
Montana	482,825	523,045	391,384	108.3	81.1
Nebraska	796,793	931,234	753,912	116.9	94.6
Nevada	1,173,814	1,331,766	1,073,041	113.5	91.4
New Hampshire	614,754	676,146	537,997	110.0	87.5
New Jersey	3,553,562	3,997,308	3,316,124	112.5	93.3
New Mexico	901,388	973,685	738,415	108.0	81.9
New York	8,108,103	8,679,561	7,135,118	107.0	88.0
North Carolina	4,327,528	5,106,116	3,985,836	118.0	92.1
North Dakota	317,498	349,441	280,019	110.1	88.2
Ohio	5,127,508	5,843,980	4,928,956	114.0	96.1
Oklahoma	1,664,378	1,936,908	1,465,664	116.4	88.1
Oregon	1,675,562	1,907,428	1,585,086	113.8	94.6
Pennsylvania	5,567,315	6,289,674	5,095,426	113.0	91.5
Rhode Island	463,388	509,921	429,334	110.0	92.7
South Carolina	2,137,683	2,522,339	1,961,875	118.0	91.8
South Dakota	363,438	400,931	320,781	110.3	88.3
Tennessee	2,812,133	3,336,464	2,647,011	118.6	94.1
Texas	9,977,436	11,800,449	9,219,315	118.3	92.4
Utah	979,709	1,103,249	885,863	112.6	90.4
Vermont	322,539	344,442	257,772	106.8	79.9
Virginia	3,364,939	3,833,196	3,177,164	113.9	94.4
Washington	2,885,677	3,270,218	2,711,528	113.3	94.0
West Virginia	881,917	916,389	641,667	103.9	72.8
Wisconsin	2,624,358	2,894,197	2,454,964	110.3	93.5
Wyoming	261,868	282,207	221,220	107.8	84.5

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Appendix 2. Number of Administrative Records Race Response Data that Matched to the 2010 Census

2010 Census and Administrative Records Race Response Match by Source File	White Alone	Black Alone	American Indian or Alaska Native Alone	Asian Alone	Native Hawaiian or Other Pacific Islander Alone	Some Other Race Alone	Two or More Races
Federal Files							
Previous Census Records	143,932,741	19,590,220	924,308	6,438,417	127,055	2,533,345	1,424,434
Numident	137,393,038	23,441,066	746,723	7,463,346	211,511	169,495	N/A
IHS	N/A	N/A	1,118,731	N/A	N/A	N/A	N/A
HUD CHUMS	3,453,455	456,376	6,375	86,759	2,747	N/A	2,925
HUD PIC	1,992,676	2,768,089	30,164	140,313	10,985	N/A	21,697
HUD TRACS	826,971	659,247	8,872	68,696	789	11,021	7,359
TANF	879,503	662,025	43,363	22,129	11,578	N/A	17,428
MEDB	35,061,707	4,084,997	119,988	680,144	N/A	46,696	N/A
Commercial Files							
Commercial File 1	80,838,100	4,576,674	N/A	3,212,143	8,512	9,411	N/A
Commercial File 2	104,760,397	6,023,253	49,702	3,818,328	19,317	6,865	N/A
Commercial File 3	87,472,679	7,182,423	80,738	3,306,341	13,804	13,203	N/A
Commercial File 4	36,691,939	3,514,892	24,430	1,428,767	6,666	6,654	N/A

Note: N/A indicates that data were not available for a demographic group.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Appendix 3. Number of Administrative Records Age Response Data that Matched to the 2010 Census

2010 Census and Administrative Records Age Response Match by Source File	Age							
	Total	0-2	3-17	18-24	25-44	45-64	65-74	75 and older
Federal Files								
Previous Census Records	177,391,205	112,085	22,721,169	17,981,703	48,621,085	57,993,070	16,294,058	13,668,035
Numident	247,246,597	9,437,770	50,571,159	22,276,150	63,311,618	68,127,690	18,537,623	14,984,587
IHS	1,982,375	58,605	461,447	235,454	594,965	471,707	103,237	56,960
HUD CHUMS	1,102,445	N/A	N/A	45,498	662,649	318,821	51,624	23,853
HUD PIC	5,629,036	282,150	2,138,739	576,799	1,157,731	953,821	278,559	241,237
HUD TRACS	1,933,895	117,310	458,322	184,889	292,834	303,777	241,166	335,597
SSR	5,752,842	77,311	917,048	487,295	1,130,532	1,875,155	655,365	610,136
SSS	11,362,121	N/A	N/A	9,925,993	1,436,128	N/A	N/A	N/A
TANF	1,962,999	284,341	924,045	218,901	396,722	124,740	11,724	2,526
MEDB	40,976,596	154	3,050	109,286	1,482,235	6,285,208	18,055,914	15,040,749
Commercial Files								
Commercial File 1	67,267,479	2	464	2,026,853	19,023,391	30,180,482	8,947,185	7,089,102
Commercial File 2	99,550,908	2,869	264,000	5,815,815	25,918,317	44,461,470	12,992,241	10,096,196
Commercial File 3	99,034,982	N/A	N/A	5,691,945	28,355,217	42,634,879	12,458,996	9,893,945
Commercial File 4	44,394,816	N/A	N/A	2,913,010	15,573,662	17,211,069	4,832,777	3,864,298
Commercial File 5	105,057,114	N/A	45	306,784	31,025,912	49,538,772	13,600,440	10,585,161
Commercial File 6	672,329	5	10	47,364	367,703	231,700	22,390	3,157
Commercial File 7	69,141,668	N/A	N/A	959,538	16,433,959	32,667,361	10,405,506	8,675,304
Commercial File 8	94,729,380	2	4,865	227,067	27,324,495	45,364,561	12,381,914	9,426,476

Note: N/A indicates that data were not available for a demographic group.

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Appendix 4. Number Coverage of 2010 Census Race Data by Administrative Records Source Files

Coverage of 2010 Race Data by Administrative Records Race Response Data by Source File	White Alone	Black Alone	American Indian or Alaska Native Alone	Asian Alone	Native Hawaiian or Other Pacific Islander Alone	Some Other Race Alone	Two or More Races
Federal Files							
Previous Census Records	158,912,424	22,172,600	1,622,741	7,262,106	236,058	6,240,160	4,347,896
Numident	147,538,811	26,032,144	1,584,438	9,410,029	310,889	1,273,753	3,317,079
HUD CHUMS	3,737,173	553,772	28,399	141,411	6,371	211,123	89,764
HUD PIC	2,354,609	3,236,689	83,620	172,802	20,131	495,651	356,228
HUD TRACS	980,146	798,188	22,003	86,587	2,583	109,232	87,265
IHS	184,336	10,859	1,213,533	1,525	557	12,263	243,411
MEDB	37,838,250	4,525,110	268,365	1,248,345	34,817	457,020	481,100
TANF	1,014,945	781,662	66,149	31,426	17,247	105,998	151,340
Commercial Files							
Commercial File 1	86,991,717	11,239,910	455,389	3,961,908	65,395	400,642	1,300,500
Commercial File 2	123,505,687	17,879,478	1,048,547	6,103,276	145,686	5,309,302	2,799,946
Commercial File 3	97,488,071	12,694,945	663,084	4,381,886	89,047	487,464	1,636,702
Commercial File 4	41,413,258	6,574,714	309,933	1,906,532	44,160	246,747	831,043

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.

Appendix 5. Number Coverage of 2010 Census Age Data by Administrative Records Source Files

Coverage of 2010 Age Data by Administrative Records Age Response Data by Source File	Age						
	0-2	3-17	18-24	25-44	45-64	65-74	75 and older
Federal Files							
Previous Census Records	181,917	26,175,297	21,416,592	55,633,830	65,122,878	18,253,613	15,773,075
Numident	10,755,280	56,519,662	26,024,257	71,248,022	75,612,369	20,497,398	17,357,698
HUD CHUMS	464	2,530	260,965	2,904,253	1,385,413	219,255	101,905
HUD PIC	346,868	2,526,583	708,426	1,382,735	1,147,854	336,633	298,257
HUD TRACS	146,157	557,649	229,341	358,202	370,604	288,031	405,192
IHS	70,172	537,573	282,007	690,349	546,098	120,249	69,654
MEDB	2,467	17,131	150,614	1,805,897	7,278,641	19,674,307	16,967,643
SSR	95,226	1,077,384	585,922	1,383,920	2,316,885	839,000	813,067
SSS	5,599	63,968	11,464,531	1,710,190	27,464	5,366	4,243
TANF	342,572	1,086,642	262,594	468,505	149,436	14,908	4,305
Commercial Files							
Commercial File 1	6,757	48,352	2,968,988	26,060,043	40,555,390	12,096,480	9,912,223
Commercial File 2	29,025	550,213	7,922,515	32,154,081	53,039,219	15,361,780	12,416,700
Commercial File 3	9,894	317,884	7,927,196	34,646,629	50,699,174	14,763,017	12,121,167
Commercial File 4	5,668	129,303	3,885,747	18,792,309	20,482,692	5,752,650	4,823,882
Commercial File 5	8,631	58,890	520,819	37,306,506	60,386,506	16,774,769	13,617,147
Commercial File 6	173	5,328	72,300	497,604	315,476	29,811	5,150
Commercial File 7	4,500	31,905	1,391,778	20,070,632	39,821,425	12,793,870	11,066,280
Commercial File 8	7,719	59,100	355,481	32,820,483	54,107,625	14,729,763	11,540,612

Sources: 2010 Census and 2010 Census Match Study Administrative Records Data.