2010 American Community Survey Match Study: Comparing Administrative Records and ACS Data

by

Adela Luque
Center for Administrative Records Research & Applications
U.S. Census Bureau
adela.luque@census.gov

Renuka Bhaskar
Center for Administrative Records Research & Applications
U.S. Census Bureau
renuka.bhaskar@census.gov

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

This paper has undergone review according to the Census Bureau publications. However, any views and conclusions expressed herein are those of the authors and do not necessarily represent those of the U.S. Census Bureau.

We wish to thank Amy O'Hara (CARRA), Eric Slud (CSRM), Anthony Tersine (DSSD), Mark Asiala (DSSD) and Deborah Griffin (ACSO) for their valuable comments and suggestions. We also thank Kuo M. Tung (ACSO), Donald Keathley (DSSD), Steven Hefter (DSSD), Aileen Bennet (ACSO), David Raglin (ACSO), Sonya Rastogi (CARRA), James Noon (CARRA) and Teresa A. Schellhamer (CARRA) for their helpful input.

Abstract

Using administrative records data from federal government agencies and commercial sources, the 2010 ACS Match Study measures administrative records coverage of 2010 ACS addresses, persons, and persons at addresses at different levels of geography as well as by demographic characteristics and response mode. The 2010 ACS Match Study represents a continuation of the research undertaken in the 2010 Census Match Study, the first national-level evaluation of administrative records data coverage.

Results indicate that administrative records provide substantial coverage for addresses and persons in the 2010 ACS (92.7 and 92.1 percent respectively), and less extensive though substantial coverage, for person-address pairs (74.3 percent). In addition, some variation in address, person and/or person-address coverage is found across demographic and response mode groups. This research establishes a baseline for future administrative records research using the ACS and informs potential uses of administrative records in survey and decennial census operations to address the increasing costs of data collection and declining response rates.

1. Introduction

In response to the increasing costs of providing demographic and socio-economic statistics for a country's population, many industrialized economies have developed alternative and less costly methods to conduct their censuses using existing administrative records data from population registers. For example, Austria, Denmark, Finland, the Netherlands, Sweden, and Switzerland already use administrative records in part or entirely in their censuses (Farber and Leggieri 2002, Asher 2010, Mulalic 2011, Ralphs and Tutton 2011). Other European countries, such as Spain, Czech Republic, Estonia, Italy, Latvia, and Lithuania use a combination of registers and field enumeration simultaneously (Valente, 2010).

While the costs of conducting the demographic census have also skyrocketed in the U.S. during the last two decades (U.S. GAO, 2012), administrative data have not yet been highly utilized in the Decennial Census or survey operations.[1] Government agencies and external researchers have used survey and administrative data primarily to enhance their datasets and evaluate differences in the reporting of common items.[2]

In an effort to reduce costs while preserving quality, the U.S. Census Bureau is researching options on the best use of administrative records in the Decennial Census. Options investigated range from nonresponse followup (e.g. Zanutto and Zaslavsky 2002, Judson and Bauder 2003) to an administrative records census (Farber and Leggieri 2002, Berning 2003, Bye and Judson 2004). Previous studies, in particular AREX 2000, examined the use of administrative records as a source of information for enhancing census operations.[3] However, it had some limitations, including the use of only federal sources, the restriction of its analysis on count coverage to five counties, and a lag between administrative records reference periods and Census Day (April 1).

Awareness of the need for further research culminated with the 2010 Census Match Study (CMS), the first to link administrative records to Decennial Census data to evaluate the quality and coverage of administrative records for the entire nation (Rastogi and O'Hara 2012). The 2010 CMS expanded AREX 2000 research capability by utilizing nine commercial files as well as four additional federal files, employing data that were close to the Decennial Census reference date of April 1, and by conducting a nation-level evaluation, rather than one limited to just five counties.

The 2010 CMS evaluated counts and matches of addresses, persons, and persons at addresses at different levels of geography as well as by demographic characteristics and response mode. It also assessed the quality and coverage of demographic data in administrative records relative to the 2010 Census, and in addition, evaluated the capability of administrative data to identify occupancy status and to estimate household population count.

The 2010 American Community Survey Match Study (ACS-MS) represents a continuation of the research undertaken in the 2010 CMS, and relies on the data compiled by and methodology employed in that study. This paper is the first of a series of studies that will extend the evaluation of administrative records coverage and quality beyond 2010 using 1-year ACS data. Specifically, this study evaluates administrative records coverage of 2010 ACS addresses, persons, and person-address pairs at different levels of geography as well as by demographic characteristics and response mode.

The objective of this report is not to show or evaluate administrative records coverage of the underlying population(s) that the ACS is designed to represent. Instead, its intent is to assess whether 2010 ACS records are present or absent in administrative data, and to establish a baseline for future administrative records research using the ACS. This type of baseline evaluation is necessary in order to determine whether there is sufficient administrative records coverage of ACS records to pursue the use of the ACS in future administrative records

[1] Administrative data have, however, been integrated into a number of important Census Bureau 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.
[2] For example, see Meijer et al. 2010, Iams and Dushi 2009, and Groen 2012.
[3] See Alvey and Scheuren 1982, Edmonston and Schultze 1995, Steffey and Bradburn 1994, Farber and Leggieri 2002, Berning 2003, Bye and Judson 2004, and the 2010 CMS (pages 2 and 3).

coverage research.[4] Further research will also provide initial evidence for whether administrative records can enhance ACS content beyond Decennial Census short-form items, such as household income.

2. Data and Methodology

The administrative records data used in this report were compiled in the 2010 CMS. The sections that follow briefly describe the methodology, and federal and commercial administrative records data used in this study.[5]

2.1. Record Linkage

To be able to match addresses and persons in the 2010 ACS to addresses and persons in administrative records, both data universes must contain common address and person identifiers. Unique address identifiers called master address file identification numbers (MAFIDs) and person identifiers called protected identification keys (PIKs) were assigned to administrative records. The 2010 ACS data already had MAFIDs, therefore only persons in the 2010 ACS needed to be assigned PIKs.

To assign address identifiers, administrative data were matched to an extract from the Census Bureau Master Address File (MAF),[6] and MAFIDs were then assigned to administrative records with address data that matched to the MAF. To assign PIKs to the 2010 ACS and to administrative data, a matching software (Person Validation Identification System or PVS) was used to compare Personally Identifiable Information (PII) from the 2010 ACS and administrative data to PII on person reference files.[7] Many federal administrative files contained SSNs, but the 2010 ACS and most commercial data did not include them. For these data sets, the search modules in PVS compared name, address, and date of birth fields to the person reference file. For more information on this record linkage system see Wagner and Layne (2012).

2.2. Federal Administrative Records Data from Other Agencies

Below are the eleven federal data sources used in this report.[8] It is worth noting that some data sources are meant to target coverage of specific populations. For instance, the Medicare Enrollment Database is meant to obtain better coverage for older people.

- Individual Income Tax Returns (Form 1040)
- Information Returns (Forms 1099 and W2)
- Department of Housing and Urban Development (HUD) Public and Indian Housing Information Center (PIC)
- HUD Tenant Rental Assistance Certification System (TRACS)
- HUD Computerized Homes Underwriting Management System (CHUMS)
- Social Security Administration (SSA) Supplemental Security Income Record
- Selective Service System Registration File (SSR)
- Centers for Medicare and Medicaid Services Medicare enrollee data (MEDB)
- Indian Health Service Patient Registration File (IHS)
- U.S. Postal Service National Change of Address File
- Temporary Assistance for Needy Families (TANF)

[4] Including 2020 Decennial Census research.
[5] See 2010 CMS, pages 3-8 for a more in depth discussion.
[6] The 2010 CMS 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.
[7] See the 2010 CMS.
[8] In addition to these federal data files, the Death Master File from SSA was used in the 2010 CMS to help determine whether a person in administrative records was alive as of April 1, 2010.

Individual Income Tax Returns provide data for individuals who file a 1040 tax return, 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.).

HUD PIC data are maintained by HUD for persons participating in the public housing program and other rental assistance programs. HUD TRACS contains data for persons receiving rental assistance and participating in other assisted housing programs through HUD. HUD 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 file includes address, personal identifiers, and date of birth for Supplementary Security Income (SSI) recipients. The 2010 CMS (as well as this study) primarily used 2010 SSA Supplemental Security Record file for SSI recipients and appended information on children and spouses from a separate 2011 file.

The MEDB from the Centers for Medicare and Medicaid Services contains Medicare enrollee data and name, address, date of birth, race, Hispanic origin, sex, and SSN. The SSR 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 section of this report. TANF data could not be used for the address or person-address analysis as the TANF file does not include addresses on the file.

2.3. Commercial Administrative Records Data

The U.S. Census Bureau acquired nine data files containing identifying information and demographic characteristics from five commercial data vendors (Experian, Targus, Veteran Service Group of Illinois, InfoUSA and Melissa Data Base Source).[9] Most of the files contain current information on address, name, race, Hispanic origin, age, and sex data, and a few contained historical data on the same variables. One of the files provided by Targus only contained addresses. Thus, this file was not used for the person and person-address analyses as the file does not contain person data.

2.4. 2010 ACS Data

Two 2010 ACS files were used in this study: the address *response* file and the person *response* file for the fifty states and the District of Columbia in the United States.[10] Persons living in households that were sent a questionnaire but never responded are not part of the analysis.

---

[9] Commercial data vendors are described by name in the Data & 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.
[10] Puerto Rico was not included as part of the analysis in this report.

Note that the analysis conducted in this report uses *unweighted* ACS response data. Weights are omitted because this study is *not* using the ACS to evaluate administrative records coverage of the U.S. population, but simply to assess whether addresses, persons and person-address pairs in the 2010 ACS response files are also present (or are observed) in the administrative records data at hand.

The analysis conducted in this study is a descriptive one-to-one match between records in administrative data and records in the 2010 ACS response data. Weighting the ACS data would imply in some cases a 1-to-2 or 1-to-3 match. The results of this type of exercise would have implications that deviate from the intent of this study.[11] For instance, a particular individual in CAPI (computer-assisted personal interview) that receives a weight of 3 in the ACS may be in administrative records data, but that does not imply that administrative records would necessarily have the other two individuals that the weighted individual represents. Analogously, the absence of that individual in administrative records does not imply that the other two individual he/she represents are also absent in administrative records.

The address response file consists of all housing unit and group quarter addresses that either responded to the survey or were determined to be vacant. The person response file contains all the persons living at the non-vacant addresses that responded to the survey by either mail, computer-assisted telephone interview (CATI) or computer-assisted personal interview (CAPI). The CAPI category can be further subdivided into those that were directly sampled into CAPI, and those that were included in CATI, but did not respond via CATI and were then subsampled into CAPI. Households that were directly sampled into CAPI are those with unmailable/undeliverable or Remote Alaska addresses. Table 1 in Appendix II presents the race, Hispanic origin, age and sex distributions of each mode and CAPI subcategory.

It is important to note that the characteristics and response rates of households (and persons) underlying these response modes are different from one another in a non-random manner. That is, there is self-selection into response mode, and this self-selection tends to be correlated with administrative records presence. With this in mind, selected match results by mode category and CAPI subcategories are presented and discussed in Appendices IV, V and VI of this report.

Administrative records specific to group quarters were not sought or acquired at the time of this report.[12] Thus, as shown in the Results sections, the coverage of group quarter addresses and people was quite low. Hence, much of the analysis in this report is restricted to addresses corresponding to housing units and the population living in them.

2.5. Duplicates

The same people and addresses may be present multiple times within the same administrative records data source. When this occurs, administrative records address, person and person-address pair files were unduplicated in order to match and evaluate them relative to the 2010 ACS.

The 2010 ACS also contained person records with duplicate PIKs. These duplicate PIKs were not removed or unduplicated since research[13] suggests that most duplicate PIKs cannot be decisively categorized as true duplicate person records. Instead, they are likely to be the result of assigning the same PIK to two different persons in error.

---

[11] Furthermore, future research will evaluate the quality of race and Hispanic origin administrative records data. This type of evaluation can only be done using unweighted data since the exercise necessitates linking persons in the ACS to the same persons in administrative records.

[12] A group quarter is "a place where people live or stay, in a group living arrangement, that is owned or managed by an entity or organization providing housing and/or services for the residents". Group quarters include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, and workers' dormitories. For further information on group quarters, see http://www.census.gov/acs/www/Downloads/data_documentation/GroupDefinitions/2011GQ_Definitions.pdf

[13] This research refers to an upcoming paper examining duplicate PIKs in the 2009 and 2010 ACS and CPS by A. Luque and B. Clark.

2.6. Differences in Address, Person and Person-Address Data between the 2010 ACS and Administrative Records Data

As it is known, the ACS is a rolling sample where households are interviewed and data are collected throughout the year by using independent monthly samples.[14] The survey reference date, thus, varies across households depending on when they were interviewed. While future work will be conducted to identify the living status of a person in administrative records at a given point in time, creating such indicator was beyond the scope of this study. Hence in this report, all persons in the administrative records universe were included regardless of their alive status (according to administrative records).

It is also worth noting that there are differences in the timing of the administrative data relative to the ACS. The administrative data used in this study had reference dates of April 1, 2010 or prior, whereas as already pointed out, the ACS was conducted throughout all twelve months of 2010. So, there is a time frame misalignment between the ACS and the administrative data used in this study.

In addition, there is a lag for small children in administrative records data. Tax data are an important source of information on children, but babies born in 2010 would not be claimed on 2009 taxes. As a result, they are not in administrative records – however, there are babies that were born throughout 2010 that are in the 2010 ACS. This phenomenon will be referred to as "the tax code" effect.

Administrative data sometimes have conflicting information regarding person-address pairs - as the data can have a given person living at multiple (different) addresses. For instance, one data source may have a person living at an address in New York, while another data source has the same person living in Florida. To compare administrative records to the 2010 ACS, all available unique person-address pairs in administrative records were used in the match to 2010 ACS person-address records. This is in contrast to the 2010 CMS, where a best address was chosen (as of April 1, 2010, Census Day) for persons with multiple addresses in administrative records.[15] It is beyond the scope of this study to select a best address for a given person at a given point in time during the survey year.

Another source of discrepancy between the ACS and administrative records regarding a person's residence is the ACS' residency rules. According to the ACS, anyone who is currently living or staying at the sample address for more than two months is considered a resident of that address. Therefore, the ACS may place persons at addresses that are not their usual residence (as long as they stay there for more than two months). By contrast, administrative records addresses are mailing addresses, which normally are not temporary short-term addresses.

2.7. Match Ratio Definition

The match ratio or rate is a measure meant to evaluate the coverage of administrative records data with respect to the 2010 ACS response files. It is calculated by dividing the unweighted count of 2010 ACS records that match to administrative records by the 2010 ACS record count and multiplying the result by 100. That is, it simply represents the percentage of 2010 ACS addresses, persons or person-address pairs that match to administrative records by MAFID, PIK, and PIK-MAFID, respectively.

In most tables in the person and person-address coverage results, the match ratio is expressed both as a percentage of all persons in the ACS (regardless of whether or not the person obtained a PIK), and as a percentage of only persons with PIKs.[16] Expressed this way, the match ratio provides a measure of administrative records coverage for those persons in the 2010 ACS that were "matchable" to administrative records. In this manner, the percent of persons that match is not comingled with the effect of PIK assignment rates. Throughout the document, the term 'match ratio' implicitly refers to the ratio expressed as a percentage of all persons in the 2010 ACS. Otherwise, it is explicitly stated.

---

[14] For more information about the ACS, see U.S. Census Bureau (2009).

[15] The best address for a given person was selected by a logistic model. For more information on the best-address model, see the 2010 CMS.

[16] The 2010 CMS reports match ratios only as a percentage of all persons in the 2010 Census.

This study reports general address, person and person-address coverage results for housing units and group quarters independently. This is in contrast with the 2010 CMS, which did not distinguish between housing unit and group quarter addresses, or between people living in housing units and those residing in group quarters. However, because existing administrative records coverage of group quarters is considerably lower than that of housing units, only the universe of housing units is used in the in-depth coverage analyses by demographic characteristics, geographic area and response mode.

2.8. Limitations

The 2010 ACS-MS includes validated addresses and persons in administrative data; that is, addresses with MAFIDs and persons with PIKs. Administrative records lacking complete or quality data to match to the MAF or the person reference file were omitted from the analysis. Meanwhile, although the analysis includes all ACS person records regardless of whether they have a PIK, these records are not part of the "matchable" set. Further research on improving PIK assignment is underway.

The administrative records data used in this study do not contain information on many ACS items such as tenure or relationship information. Future research will evaluate how administrative records can be used to enhanced the imputation of such items when missing.

2.9. Caveats for Comparison to 2010 Census Match Study Results

No strict comparison can be made between the results obtained in this study and those in the 2010 CMS. First and foremost, the CMS represents the universe of the U.S. population while, as mentioned earlier, this study employs *unweighted* 2010 ACS *response* data. Results presented in this study are not intended to represent the administrative records coverage of the underlying U.S. population.[17] For instance, the match ratio among Hispanics in the 2010 CMS was 77.2 percent, indicating that 77.2 percent of Hispanics in the 2010 Census population were found in administrative records. In this study, the match ratio among Hispanics was 84.3 percent. This simply indicates that 84.3 percent of Hispanics for whom a response was obtained in the 2010 ACS were present in administrative records.[18]

In addition, and as stated earlier, this study uses all unique person-address combinations available in the administrative records at hand. On the other hand, the 2010 CMS employed a best-address model to select a unique address for each individual as of April 1, 2010. In addition, the 2010 CMS applied an "alive" status indicator to persons in administrative records and only included in the analysis those persons that were alive according to the indicator on Census Day. The current study does not employ an alive flag indicator. Finally, the 2010 CMS did not present separate coverage results for housing units and group quarters. By contrast, the 2010 ACS-MS differentiates between the two types of living quarter. It presents overall coverage results separately for both, and detailed results for just housing units.

With these strong caveats, a descriptive, non-statistical comparison to the main 2010 CMS results is presented in this report. Such a comparison adds value to this study since it will help identify potential red flags in administrative records coverage of the ACS, and serve as a baseline for future ACS-administrative records research throughout the decade.[19] For instance, the 2010 CMS found that administrative records coverage among older age groups is higher

[17] Weights in the ACS are meant to produce representative estimates of the underlying population. However, it is not clear whether they would produce representative estimates of the underlying population that matches to administrative records. This is because the matching subset may be different from the ACS sample in a non-random manner. It could also be argued that this non-randomness may be exacerbated for the matching subset of person-address pairs.

[18] Nevertheless, as indicated earlier, there is self-selection into response mode, and this self-selection tends to be correlated with administrative records presence. To shed some light on this issue, selected match results by mode category and CAPI subcategories are presented and discussed in Appendices IV, V and VI of this report.

[19] Changes to the ACS sample will be taken into consideration as administrative records coverage research progresses throughout the decade –as these changes may impact administrative records coverage of the ACS overtime.

than among younger ones. If this study revealed otherwise, a closer examination would have to be undertaken. In some cases, a drill-down analysis was conducted to shed light into the administrative records coverage behavior of a particular type of address or demographic group.[20]

3. Address Count and Match Results

This section presents results on the number and percent of ACS addresses that match to administrative records by various housing, geographic, household demographic characteristics, and response mode subcategories. Match results among demographic groups by mode subcategories are presented and discussed in Appendix IV. This section also discusses results in relation to those found in the 2010 CMS.

The match counts and ratios that follow represent the match of the 2010 ACS to federal and commercial administrative records. However, in order to assess the contribution of commercial data to administrative records coverage, the state table presents state-level match ratios for federal and commercial data separately. As discussed earlier, all the numbers and percentages presented below do not include any weights.

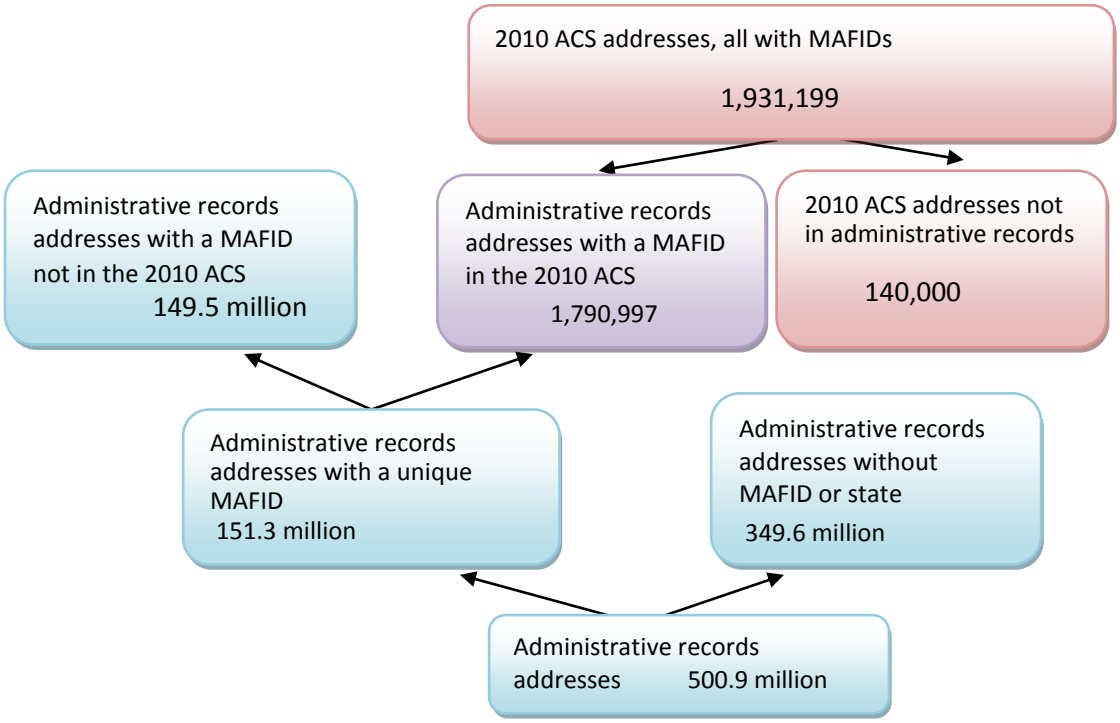
3.1. Overall

Figure 1 below shows the number of addresses in the 2010 ACS response file and administrative records. As already mentioned in the section 2.1, unique identifiers called MAFIDs are assigned to addresses and are used to facilitate record linkages between the 2010 ACS and administrative records.
There were 1.9 million addresses in the 2010 ACS response file, and all of these address had MAFIDs. There were 151.3 million addresses in administrative records that had a unique MAFID and 349.6 million addresses that did not have a MAFID.[21] As expected, there were more addresses in the administrative records relative to the 2010 ACS response file since the ACS is a sample dataset.

Of the 1.9 million addresses in the 2010 ACS response file, 1.8 million (or 92.7 percent) matched to MAFIDs in administrative records. There were 140,000 addresses that were in the ACS but were not in administrative records. Definitional differences between addresses in the 2010 ACS and administrative records contributed to the address non-matches. For example, the 2010 ACS included some physical descriptions of addresses such as “yellow house near fork in the road,” which cannot be matched to administrative data since administrative records only contain mailing addresses.

[20] As mentioned earlier, key match results are presented for demographic-mode groups (including CAPI subcategories) in order to better assess their administrative records coverage behavior, which tends to be correlated with response mode.
[21] Current research is examining the reasons for these addresses not obtaining a MAFID and exploring ways to increase MAFID assignment.

Figure 1. Count and Match of 2010 ACS and Administrative Records Addresses
Universe: All addresses



Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

3.2. Housing Units and Group Quarters

The 2010 ACS includes both housing units and group quarters. Most of the MAFIDs in the 2010 ACS represented housing units, thus the match ratios for all MAFIDs are similar to those of the housing units. As shown in Table 1, of the 1,917,799 million housing unit MAFIDs, administrative records matched to 1,783,705 million or 93.0 percent. The match ratio was notably lower for group quarters at 54.4 percent, implying that administrative records coverage of group quarters addresses is lower than the coverage of housing unit addresses.

Table 1. 2010 ACS and Administrative Records Address Match by Type of Living Quarter
Universe: All addresses

ACS Type of Living Quarter	2010 ACS Address Count	Match of 2010 ACS and Administrative Records Addresses	Match Ratio 2010 ACS and Administrative Records Addresses
Total	1,931,199	1,790,997	92.7
Housing Units	1,917,799	1,783,705	93.0
Group Quarters	13,400	7,292	54.4

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

As mentioned earlier, the majority of federal and commercial administrative records data used in this analysis did not include records specific to group quarters. Because of the low coverage for group quarters, the remainder of this section focuses on housing unit addresses only.

3.3. Region

Table 2 below shows the number of 2010 ACS addresses as well as the administrative records and 2010 address match numbers and ratios by region.

Table 2. 2010 ACS Address Count and Administrative Records Match Counts and Ratios by Region
Universe: Housing Units

ACS Region	2010 ACS Address Count	2010 ACS Percent	2010 ACS and Administrative Records Address Match	2010 ACS and Administrative Records Address Match Ratio
Total	1,917,799	100.0	1,783,705	93.0
Northeast	362,261	18.9	327,486	90.4
Midwest	524,496	27.3	494,941	94.4
South	648,389	33.8	604,802	93.3
West	382,653	20.0	356,476	93.2

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

All regions had a match ratio higher than 90.0 percent. The Midwest had the highest match ratio with 94.4 percent, the South had the next highest match ratio with 93.3 percent followed closely by the West (93.2 percent) and finally, the Northeast (90.4 percent).

3.4. State by Federal and Commercial Administrative Records Data

Table 3 shows the number of addresses in the 2010 ACS and the match numbers and ratios by state. It also highlights the contribution of commercial data to overall and state-level match ratios. Specifically, it presents state-level match ratios for federal administrative records (fourth column) as well as for combined federal and commercial administrative records (fifth column). Finally, the last column shows the increase in the match ratios attributable to commercial data.

Overall, over 85 percent of addresses in the 2010 ACS matched to unique MAFIDs in federal administrative data (fourth column). Incorporating commercial data resulted in an increase in the overall address match ratio of 7.6 percentage points.

At the state level, federal data match ratios ranged from 52.3 percent in Alaska to 92.4 percent in Maryland. The inclusion of commercial administrative records resulted in increases in the match ratios for all states, though the level of the impact varied across them. States for which commercial data had the highest impact included Wyoming, Montana, Hawaii, Arkansas, and Delaware. For each of these states, the match ratios for addresses in the 2010 ACS increased by more than ten percentage points when commercial data was used in addition to federal records. States for which commercial data had the least impact included Maryland, Ohio, California, Virginia, Pennsylvania, and Connecticut. For each of these states, the match ratios for addresses in the 2010 ACS increased by less than six percentage points when commercial data was used in addition to federal records. These states had relatively high match ratios when only federal data were used.[22]

The state-level match ratios using the combined federal and commercial administrative records ranged from 60.8 percent for Alaska to 97.4 percent for Maryland. For most states, over 90 percent of the addresses in the ACS were

[22] For further information on the characteristics of commercial versus federal administrative records, see Tables 1 and 2 in Appendix I. These tables show the distribution of selected characteristics for addresses that matched to federal administrative records and for addresses that matched to commercial administrative records but did not match to federal administrative records. Addresses that matched to commercial data only had a higher percentage of rural addresses and lower proportions of single-family homes and householders responding via mail compared to addresses that matched to federal data. Additional research is planned to further evaluate differences in match ratios using federal and commercial data sets.

found in administrative records. The four states with the highest match ratios were in the Midwest and South, including Maryland (97.4 percent), Ohio, the District of Columbia, and Iowa, each with a match ratio of 97.0 percent. California, the state with the largest number of addresses, had the fifth highest address match ratio of 96.6 percent.

The five states with the lowest address match ratios were in the West and Northeast, contained many non city-style addresses, CAPI Direct (or unmailable/undeliverable) addresses, and generally had a relatively low number of total addresses in the unweighted data. Alaska had the lowest address match ratio at 60.8 percent followed, at some distance, by Maine (74.7 percent), West Virginia (75.5 percent), Montana (77.1 percent), and Vermont (77.6percent).

One factor that may be related to Alaska's low match ratio is that data for over half of the addresses in this state were collected through computer-assisted personal interview (CAPI), and in particular, almost 40 percent were under the CAPI Direct category. This is a much higher percent than any other state (see Appendix I, Table 3). Additionally, nearly half of the addresses in Alaska in the 2010 ACS were in rural areas - a higher percent than most states. As shown in Appendix I, Table 4, 2010 ACS CAPI addresses in rural areas were generally associated with considerably lower match ratios. In particular, these addresses had a match ratio that was at least 26 percentage points lower than any other combination of mode-urban status. Another characteristic associated with low match ratios is occupancy status – vacant housing units have a lower match ratio relative to occupied housing units (see Appendix I, Table 5). Over 85 percent of vacant addresses were CAPI (142,511 of 166,940), indicating that an in-person visit may have been necessary to determine the occupancy status of the address. Maine and Alaska had the highest percentages of vacant housing units among states in the 2010 ACS – 24.3 and 20.4 percent respectively, and vacant housing units in these states had very low match ratios (see Appendix I, Tables 6 and 7).

Table 3. 2010 ACS Address Count and Administrative Records Match Counts and Ratios by State
Universe: Housing Units

ACS State	2010 ACS Address Count	2010 ACS Addresses that match to Federal Administrative Records	Match Ratio for Federal Records	2010 ACS Addresses that match to Federal or Commercial Administrative Records	Match Ratio for Federal and Commercial Administrative Records	Percentage Point Increase in Match Ratio Attributable to Commercial Data
Total	1,917,799	1,638,811	85.5	1,783,705	93.0	7.6
Alabama	31,552	26,500	84.0	29,240	92.7	8.7
Alaska	5,604	2,929	52.3	3,407	60.8	8.5
Arizona	33,254	27,462	82.6	29,825	89.7	7.1
Arkansas	19,178	15,162	79.1	17,352	90.5	11.4
California	176,551	160,621	91.0	170,598	96.6	5.7
Colorado	30,885	26,148	84.7	28,849	93.4	8.7
Connecticut	20,503	18,503	90.2	19,722	96.2	5.9
Delaware	6,339	5,194	81.9	5,909	93.2	11.3
District of Col.	3,819	3,422	89.6	3,705	97.0	7.4
Florida	100,820	88,842	88.1	97,308	96.5	8.4
Georgia	49,876	42,938	86.1	46,741	93.7	7.6
Hawaii	7,537	5,693	75.5	6,568	87.1	11.6
Idaho	10,126	7,952	78.5	9,053	89.4	10.9
Illinois	80,592	70,990	88.1	76,603	95.1	7.0
Indiana	42,080	38,042	90.4	40,616	96.5	6.1
Iowa	28,130	25,154	89.4	27,273	97.0	7.5
Kansas	22,065	19,090	86.5	20,963	95.0	8.5
Kentucky	28,208	24,130	85.5	26,053	92.4	6.8
Louisiana	27,108	22,801	84.1	25,428	93.8	9.7
Maine	15,212	9,779	64.3	11,363	74.7	10.4
Maryland	30,811	28,469	92.4	29,998	97.4	5.0
Massachusetts	37,494	33,080	88.2	35,708	95.2	7.0
Michigan	82,647	71,111	86.0	77,491	93.8	7.7
Minnesota	56,799	47,874	84.3	52,479	92.4	8.1
Mississippi	16,790	13,332	79.4	15,145	90.2	10.8
Missouri	43,440	36,954	85.1	40,044	92.2	7.1
Montana	9,256	5,881	63.5	7,133	77.1	13.5
Nebraska	16,768	13,953	83.2	15,572	92.9	9.7
Nevada	13,310	11,703	87.9	12,610	94.7	6.8
New	10,013	7,873	78.6	8,677	86.7	8.0
New Jersey	48,420	42,665	88.1	46,080	95.2	7.1
New Mexico	12,779	8,632	67.5	9,975	78.1	10.5
New York	116,705	95,039	81.4	103,513	88.7	7.3
North Carolina	56,018	46,736	83.4	52,238	93.3	9.8
North Dakota	7,834	6,012	76.7	6,757	86.3	9.5
Ohio	76,268	69,808	91.5	73,947	97.0	5.4
Oklahoma	28,727	22,082	76.9	24,723	86.1	9.2
Oregon	23,620	20,516	86.9	22,530	95.4	8.5
Pennsylvania	100,139	84,849	84.7	90,756	90.6	5.9
Rhode Island	5,942	5,157	86.8	5,587	94.0	7.2
South Carolina	26,571	22,425	84.4	24,879	93.6	9.2
South Dakota	7,935	6,170	77.8	6,945	87.5	9.8
Tennessee	37,220	32,841	88.2	35,465	95.3	7.0
Texas	127,757	107,705	84.3	118,379	92.7	8.4
Utah	14,934	12,710	85.1	13,705	91.8	6.7
Vermont	7,833	5,249	67.0	6,080	77.6	10.6
Virginia	44,284	39,610	89.4	42,192	95.3	5.8
Washington	40,883	35,772	87.5	38,859	95.0	7.6
West Virginia	13,311	9,000	67.6	10,047	75.5	7.9
Wisconsin	59,938	51,528	86.0	56,251	93.8	7.9
Wyoming	3,914	2,723	69.6	3,364	85.9	16.4

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

3.5. Housing Unit Type

Table 4 below shows 2010 ACS and administrative record address counts and match ratios by housing unit type. Buildings with twenty or more apartments had the highest match ratio at 95.4 percent, followed closely by buildings with ten to nineteen apartments (95.0 percent), buildings with five to nine apartments (94.2 percent) and single-family homes (94.1 percent) - which make up the largest portion of all housing units. Buildings with two to four apartments had a lower match ratio (87.4 percent) relative to other multi-unit buildings. This lower match rate 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 have mail delivered to one box (Virgile 2012). Mobile homes and other housing unit types (such as boats, RVs, and vans) had the lowest match ratios with 82.9 and 73.1 respectively.

Table 4. 2010 ACS Address Count and Administrative Records Match Counts and Ratio by Housing Unit Type

Universe: Housing Units

ACS Housing Unit Type	2010 ACS Address Count	Match of 2010 ACS and Administrative Records Addresses	Match Ratio 2010 ACS and Administrative Records Addresses
Total	1,917,799	1,783,705	93.0
Building with 2-4 apartments	128,800	112,549	87.4
Building with 5 to 9 apartments	69,606	65,590	94.2
Building with 10 to 19 apartments	61,870	58,802	95.0
Building with 20+ apartments	131,946	125,829	95.4
Other (Boat/RV/van etc.)	1,467	1,073	73.1
Single-family home	1,398,315	1,315,623	94.1
Mobile home	125,795	104,239	82.9

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

3.6. Race and Hispanic origin of Householder and Mode of Collection

Thus far, the figures and tables have focused on the 1.9 million housing unit addresses in the 2010 ACS, regardless of whether they were occupied or vacant. However, the universe for this sub-section is occupied housing units since it examines match ratios by the demographic characteristics and response mode of the householder.

Table 5 shows 2010 ACS match ratios by ACS Hispanic origin and race of the householder as well as by mode of data collection. The CAPI category is further divided into two separate subcategories in order to explore their match ratio differences: CAPI addresses with an unmailable or undeliverable address that were directly subsampled into CAPI (CAPI direct), and CAPI addresses that were mailed a questionnaire but have not responded (CAPI nonresponse).[23]

Of these occupied housing units, administrative records matched to 1.7 million or 95.2 percent.[24] Address match ratios for non-Hispanic and Hispanic householders were high and virtually identical (95.2 percent and 95.0 percent for non-Hispanics and Hispanics respectively). This evidence indicates that administrative records provide similar address coverage for both Hispanic and non-Hispanic headed households.

[23] See http://www.census.gov/acs/www/Downloads/survey_methodology/Chapter_4_RevisedDec2010.pdf.

[24] Administrative records have a higher match ratio for occupied housing units relative to vacant housing units due to the nature of federal and commercial data gathered for the project. For more information on differences in match ratios occupancy status, see Tables 3 and 5 in Appendix I.

With the exception of the American Indian or Alaska Native (AIAN) group, match ratios for all race groups were above 90 percent. The highest match ratio (96.5 percent) was for housing units where the head of household was Asian alone. The match ratio was much lower for addresses where the head of household was AIAN alone at 77.9 percent.

The bottom rows of Table 5 show the match ratios by response mode. There was virtually no difference between the match ratios of households responding by mail (97.2 percent) and via computer-assisted telephone interview (CATI) (97.1 percent). By contrast, households responding via CAPI had a match ratio nearly 15 percentage points lower (at 85.4 percent) relative to those responding by mail. This lower match ratio for CAPI households can be explained in part by the fact that incomplete addresses or addresses that are physical descriptions of a location (i.e., unmailable or undeliverable), and addresses in Remote Alaska are sampled and sent directly to CAPI.[25] When CAPI addresses are broken into unmailable/undeliverable and non-response addresses, the match ratio for the unmailable/undeliverable addresses is quite low (28.9 percent) as expected. The match ratio for the remainder of CAPI addresses, for which a questionnaire was initially mailed but no response was received by mail, is 91.7 percent. This ratio is much higher and closer to the ratios found for mail and CATI households. For a break-down of address match ratios for demographic groups by mode categories and a discussion of results, see Appendix IV.

Table 5. 2010 ACS Address Count and Administrative Records Match Counts and Ratios by Race and Hispanic Origin of Householder and Mode

Universe: Occupied housing units

ACS Demographic Characteristics of Householder and Mode	2010 ACS Address Count	Match of 2010 ACS and Administrative Records Addresses	Match Ratio 2010 ACS and Administrative Records Addresses
Total Housing Units	1,750,859	1,666,336	95.2
Hispanic or Latino Origin			
Hispanic	153,640	145,992	95.0
Non Hispanic	1,597,219	1,520,344	95.2
Race			
White Alone	1,438,326	1,370,302	95.3
Black Alone	165,126	157,878	95.6
American Indian or Alaska Native Alone	14,649	11,414	77.9
Asian Alone	62,106	59,954	96.5
Native Hawaiian and Other Pacific Islander Alone	1,677	1,592	94.9
Some Other Race Alone	40,922	38,769	94.7
Two or More Races	28,053	26,427	94.2
Mode			
Mail	1,251,082	1,215,906	97.2
CATI	201,262	195,413	97.1
CAPI	298,515	255,017	85.4
CAPI Direct	29,615	8,570	28.9
CAPI Nonresponse	268,900	246,447	91.7

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

[25] See http://www.census.gov/acs/www/Downloads/survey_methodology/Chapter_4_RevisedDec2010.pdf.

3.7. 2010 Census Match Study

As discussed earlier in Section 2.9, although results from the 2010 ACS cannot be strictly compared to those from the 2010 CMS, a descriptive, non-statistical comparison is valuable to round off the overall assessment of administrative records coverage with respect to the ACS. It is with the strong caveats discussed in section 2.9 that this subsection discusses the previous address results in the context of the 2010 CMS. Recall that the 2010 ACS data are unweighted and results reflect the presence or absence of administrative records for the addresses in the survey response file. The 2010 ACS results are not intended to reflect all housing units or the entire U.S. address population.

Contrasting region and state results between the two studies reveals no red flags. Instead, both sets of results show a similar pattern of match ratios by region and state.[26] The five states with the highest address match ratios in the 2010 ACS were also among the top ten states in terms of match ratios in the 2010 Census. Similarly, the five states with the lowest match ratios for the ACS were the same five states with the lowest match ratios for the 2010 Census.

The classification of housing unit type used in this analysis differs from that used in the 2010 CMS. Therefore, the results are not comparable. In this analysis, housing unit type was based on responses to an ACS question on the description of the building.[27] In the 2010 CMS, housing unit type was classified based on information from the Master Address File (see 2010 CMS, page 21). Nevertheless, the match ratios described here followed a similar pattern to match ratios for addresses in the 2010 Census. For both the 2010 Census and the 2010 ACS, administrative records had the lowest match ratio for addresses classified as "Other" (including boats, RVs, vans, etc.). Additionally, both analyses show that administrative records had a lower match ratio for buildings with two to four apartments or units relative to multi-unit buildings with a larger number of units and to single-family homes.[28]

The 2010 CMS found a similar pattern for address match ratios by Hispanic origin and race of the householder. Match ratios for the 2010 Census by Hispanic origin were close in magnitude but slightly lower than those found in the 2010 ACS. Regarding race groups, match ratios in both studies were above 90 percent - with the exception of the AIAN alone group. For most race groups, match ratios were higher for the 2010 ACS relative to match ratios for the 2010 Census. However, this was not the case for households where the household head was AIAN alone.[29]

Finally, the match ratios for mail addresses in the 2010 ACS were the same as those found for mail addresses in the 2010 Census.

4. Person Count and Match Results

This section discusses administrative records coverage of persons in the 2010 ACS and presents coverage results by type of living quarter, demographic characteristics, response mode and geographic area. It also assesses the contribution of commercial administrative records data to the coverage of persons in the 2010 ACS, and discusses results in relation to those found in the 2010 CMS . Match results for demographic groups by mode subcategories are presented and discussed in Appendix V.

---

[26] The 2010 Census and unweighted 2010 ACS distribution of address are slightly different. The unweighted 2010 ACS data had a lower proportion of addresses in the South and West, and a higher proportion of addresses in the Midwest. However, the regional distribution of the weighted 2010 ACS looks very similar to the 2010 Census.

[27] The ACS question on housing unit type asks "Which best describes this building? Include all apartments, flats, etc., even if vacant."

[28] When using the same housing unit type classification as was used in the CMS, the match ratios were closer in magnitude – within 4 percentage points for all housing unit types except those classified as "Other". Once again, addresses classified as "Other" in the 2010 ACS had a much higher match ratio compared to addresses with the same classification in the 2010 Census.

[29] For these households, the 2010 Census match ratio was over 4 percentage points higher than the match ratio in the 2010 ACS. In the 2010 Census, there were enumeration efforts made in the Type of Enumeration Areas (TEAs) associated with the AIAN population. More detail on the Update Enumerate, Remote Update Enumerate, and Remote Alaska TEAs can be found in Fallica et al. (2012). The delineation of TEAs in the 2010 Census is described in Johanson et al. (2011).

The match numbers and ratios in the tables that follow represent the match of the 2010 ACS to federal and commercial administrative records. However, in order to assess the contribution of commercial data to administrative records coverage, the state table presents state-level match ratios for federal and commercial data separately.

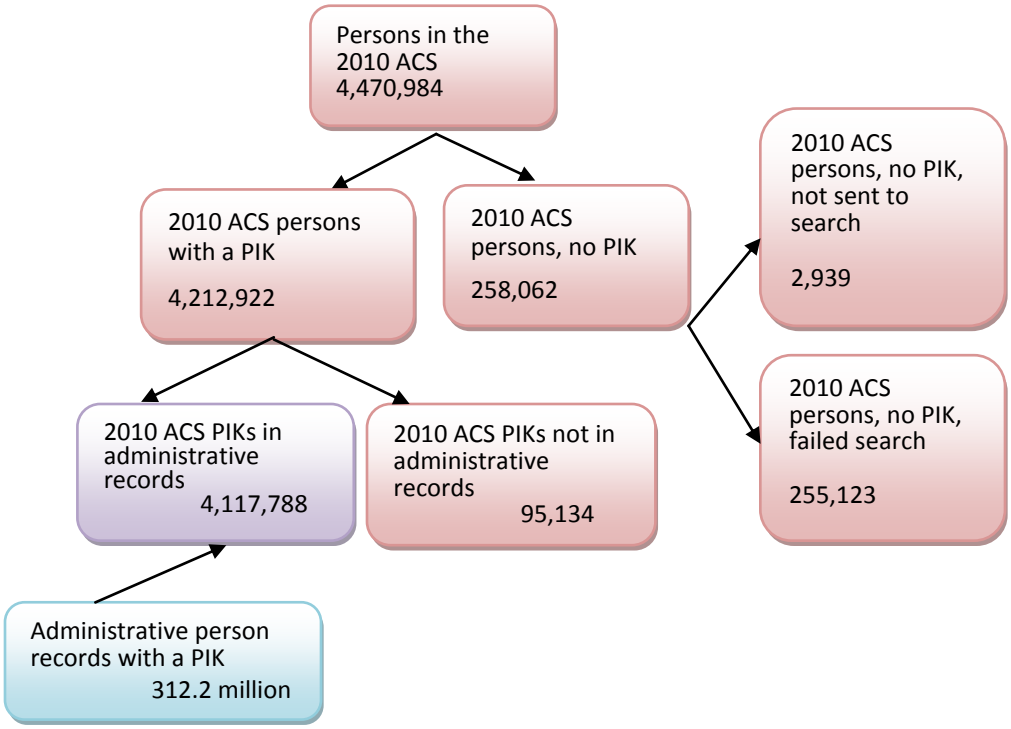
It should also be noted that person match ratios should be interpreted slightly differently relative to those for addresses. All 2010 ACS addresses had MAFIDs, therefore the entire sample of ACS addresses had the potential to be matched to administrative records with MAFIDs. By contrast, not all persons in the 2010 ACS received a PIK, and therefore, these person records did not have the potential of being matched to administrative records.[30] For this reason, match ratios are expressed as a percentage of all persons in the 2010 ACS (regardless of whether or not they received a PIK) and also as a percentage of only persons with PIKs.[31]

Analogously to the address section, this section first presents general results for all living quarters (housing units and group quarters) followed by more detailed results for just housing units. All numbers and percentages presented are unweighted.

4.1. Overall

Figure 2 shows the number of persons in the 2010 ACS and administrative records, the number of persons records in the 2010 ACS that matched and did not match to administrative records, and for those that did not match, the number that could not be processed through PVS because they lacked the necessary personal identifiable information.

Figure 2. Count and Match of 2010 ACS and Administrative Records Persons
Universe: All Persons



Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data.

[30] Characteristics of person records with PIKs and without PIKs are contrasted later in the report.
[31] As also stated in the 2010 CMS, future research will explore other matching methods, such as direct person matching across files, that do not require prior validation against a reference file.

There were 4.5 million persons in the 2010 ACS data and 4.2million (or 94.2 percent) of them obtained PIKs. Administrative records matched to 4.1 million persons in the 2010 ACS - this represented 97.7percent of ACS individuals with PIKs and 92.1 percent of all persons in the ACS (whether they were assigned a PIK or not).

Of the 258,062 persons that did not received a PIK, 2,939 (or 1.1 percent) could not be sent through the PVS process as they lacked name and date of birth, and 255,123 went through the PVS process but failed the search modules. For more information on the PVS process, see Wagner and Layne (2012).

4.2. Housing Units and Group Quarters

Table 6 highlights differences in PIK assignment rates and match ratios between persons living in housing units and group quarters in the 2010 ACS. The second column shows the count of persons in the 2010 ACS, the next set of columns displays the count and percentage of persons with PIKs, and the last set of columns shows the number and percentage of 2010 ACS persons matched to administrative records. Persons in the 2010 ACS living in housing units had a higher match ratio (92.4 percent) than those living in group quarters (83.3 percent) – a 9.1 percentage point difference. This was partly due to the fact that persons living in housing units were PIKed at a higher rate (94.4 percent) than those living in group quarters (89.2 percent). When match ratios were expressed as a percentage of persons with PIKs in the 2010 ACS, the difference between those living in housing units and group quarters was reduced to 4.5 percentage points - 97.9 percent and 93.4 percent for those in housing units and group quarters respectively.

Table 6. 2010 ACS and Administrative Records Person Match by Type of Living Quarter
Universe: All persons in the 2010 ACS

ACS Type of Living Quarter	2010 ACS Persons	2010 ACS with PIKs		Match of 2010 ACS and Administrative Records		
	Count	Count	Percent	Count	Match Ratio As a % of all persons in 2010 ACS	Match Ratio As a % of persons with PIKs in 2010 ACS
Total	4,470,984	4,212,922	94.2	4,117,788	92.1	97.7
Housing Units	4,326,036	4,083,685	94.4	3,997,069	92.4	97.9
GQs	144,948	129,237	89.2	120,719	83.3	93.4

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Analogously to the address section, the remainder of this section presents results only for persons living in housing units.

4.3. Characteristics of Person Records with PIKs and without PIKs

Table 7 contrasts the ACS demographic characteristics, region and response mode of persons with PIKs and those without PIKs living in housing units in the 2010 ACS.

The most salient result in Table 7 is that nearly 80 percent of persons without PIKs in the 2010 ACS came from households that were interviewed either by CATI or CAPI – while CATI/CAPI households comprised only 31.3 percent of the entire (unweighted) 2010 ACS response data sample.[32] The low PIK assignment rate for CATI and suggests that, relative to mail respondents, name and date of birth data collected from CATI/CAPI householders may be less reliable or absent, thus affecting PIK assignment. In fact, this finding is consistent with how some CATI/CAPI interviews may be conducted. When the respondent raises privacy concerns about names or other personal identifiable information, and in order to obtain the interview, ACS interviewers are trained to allow answers such as initials, “lady of the household”, or “gentleman of the household”.

[32] The share of person records without PIKs was over-represented in both CAPI subcategories.

In addition, further examination showed that relative to the race, Hispanic origin and age distributions of mail respondents, the distributions of CATI and CAPI Nonresponse households were skewed towards the Hispanic and Black Alone groups (see Appendix II, Table 1). CAPI Nonresponse households also had higher proportions of younger people (18-44), movers and SOR alone compared to mail respondents. Meanwhile, relative to mail, the CAPI Direct category was skewed towards the AIAN alone and rural groups. These patterns are reflected in the age, race and Hispanic origin distributions of the general unPIKed relative to the PIKed group in Table 7 below – as most of the unPIKed group is composed of CATI/CAPI respondents.[33,34]

[33] The same qualitative results were found by NORC (2011). However, their report did not examine PIK assignment patterns by response mode. Also see http://www.census.gov/acs/www/Downloads/library/2008/2008_Joshipura_01.pdf.
[34] Future research will explore the relationship between demographic, socio-economic, mode, regional and local characteristics and the likelihood of receiving a PIK.

Table 7. Distributions of Demographic Characteristics, Response Mode and Region of Persons with and without PIKs in the 2010 ACS

Universe: 2010 ACS persons living in housing units

ACS Demographic Characteristics, Response Mode, and Region	2010 ACS Persons		2010 ACS Persons with PIKs		2010 ACS Persons without PIKs	
	Count	Percent	Count	Percent	Count	Percent
Total	4,326,036	100.0	4,083,685	100.0	242,351	100.0
Hispanic or Latino Origin						
Hispanic	542,890	12.5	484,543	11.9	58,347	24.1
Non Hispanic	3,783,146	87.5	3,599,142	88.1	184,004	75.9
Race						
White Alone	3,418,850	79.0	3,248,926	79.6	169,924	70.1
Black Alone	412,411	9.5	383,547	9.4	28,864	11.9
American Indian or Alaska Native Alone	41,886	1.0	37,907	0.9	3,979	1.6
Asian Alone	189,241	4.4	175,805	4.3	13,436	5.5
Native Hawaiian or Other Pacific Islander Alone	5,919	0.1	5,317	0.1	602	0.2
Some Other Race Alone	146,981	3.4	127,792	3.1	19,189	7.9
Two or More Races	110,748	2.6	104,391	2.6	6,357	2.6
Age						
0-2	152,665	3.5	142,731	3.5	9,934	4.1
3-17	848,408	19.6	795,590	19.5	52,818	21.8
18-24	336,000	7.8	311,046	7.6	24,954	10.3
25-44	1,045,017	24.2	980,657	24.0	64,360	26.6
45-64	1,257,580	29.1	1,197,505	29.3	60,075	24.8
65-74	376,531	8.7	359,589	8.8	16,942	7.0
75+	309,835	7.2	296,567	7.3	13,268	5.5
Sex						
Male	2,091,391	48.3	1,971,169	48.3	120,222	49.6
Female	2,234,645	51.7	2,112,516	51.7	122,129	50.4
Mode						
Mail	2,972,654	68.7	2,922,566	71.6	50,088	20.7
CATI	545,695	12.6	455,290	11.1	90,405	37.3
CAPI	807,687	18.7	705,829	17.3	101,858	42.0
CAPI Direct	74,570	1.7	63,900	1.6	10,670	4.4
CAPI Nonresponse	733,117	17.0	641,929	15.7	91,188	37.6
Region						
Northeast	808,598	18.7	763,215	18.7	45,383	18.7
Midwest	1,162,122	26.9	1,114,060	27.3	48,062	19.8
South	1,433,145	33.1	1,349,763	33.1	83,382	34.4
West	922,171	21.3	856,647	21.0	65,524	27.1

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

4.4. Demographic Characteristics and Mode Match Results

Table 8 shows the number and percentage of person records with PIKs and match ratios by demographic characteristics and response mode of persons in the 2010 ACS.[35] Note that the demographic characteristics and mode classification in this table are based on data from the 2010 ACS. Match results for demographic groups by mode subcategories are presented and discussed in Appendix V.

The two more salient results in Table 8 involve the youngest age group, CATI/CAPI respondents and the SOR alone group. The youngest (0 to 2) group had a considerably lower match ratio than any other demographic group, even if only those with PIKs were taken into consideration. At the same time, the incidence of persons without PIKs was higher among CATI and CAPI respondents, and the SOR alone group relative to other groups. These findings, as well as others, are described in more detail below.

While the percentage of persons with PIKs did not vary considerably across age groups, the match ratios ranged from 75.8 percent (ages 0 to 2) to 95.6 percent (ages 75 and above) – an almost 20 percentage point difference. In the 0 to 2 age group, 93.5 percent had a PIK, but a much lower percent (75.8) were in both the 2010 ACS and administrative records. That is, almost 19 percent of those with a PIK in this age group could not be found in administrative records (data not shown).

Those aged 3 to 17 also had a lower match ratio compared to older age groups, but were a distant second relative to the 0 to 2 age group. Out of the 3 to 17 age group, 4.3 percent had a PIK but were not found in administrative records compared to 1.9 percent or less for those 18 and over (data not shown).

There are several reasons why the younger age groups, and particularly children of ages 0 to 2, had lower coverage in administrative records. As described in the 2010 CMS, "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. Babies born on or after January 1, 2010 would not be claimed on 2009 taxes, therefore they may appear in the 2010 ACS, but they would not be in the administrative records data used for this report. Additionally, tax forms such as 1040EZ do not collect data on dependents, and the IRS 1040 data used in this study only had information on the first four dependents on a tax return, potentially limiting the number of children reported in larger households."[36] Future research will assess other types of tax return data that include all dependents.

Furthermore, non-tax administrative records containing information on children had reference dates as of April 1, 2010 or prior. Therefore, 2010 ACS babies born in 2010 after Census Day effectively had zero probability of being found in administrative records.

[35] Future research will employ regression methods to explore relationships between demographic, geographic and socio-economic characteristics and the likelihood of being found in administrative records.

[36] See page 26 of 2010 CMS.

Table 8. 2010 ACS and Administrative Records Person Count and Match by Demographic Characteristics and Mode

Universe: 2010 ACS persons living in housing units

ACS Demographic Characteristics and Mode	Persons in the 2010 ACS	2010 ACS persons with PIKs		Match of 2010 ACS and Administrative Records	
	Count	Count	Percent	Match Ratio as a percentage of all ACS persons	Match Ratio as a percentage of ACS persons with PIKs
Total	4,326,036	4,083,685	94.4	92.4	97.9
Hispanic or Latino Origin					
Hispanic	542,890	484,543	89.3	84.3	94.5
Non Hispanic	3,783,146	3,599,142	95.1	93.6	98.3
Race					
White Alone	3,418,850	3,248,926	95.0	93.3	98.2
Black Alone	412,411	383,547	93.0	91.0	97.9
American Indian or Alaska Native Alone	41,886	37,907	90.5	88.8	98.1
Asian Alone	189,241	175,805	92.9	89.7	96.5
Native Hawaiian or Other Pacific Islander Alone	5,919	5,317	89.8	86.5	96.2
Some Other Race Alone	146,981	127,792	86.9	81.4	93.6
Two or More Races	110,748	104,391	94.3	90.3	95.8
Age					
0-2	152,665	142,731	93.5	75.8	81.0
3-17	848,408	795,590	93.8	89.5	95.5
18-24	336,000	311,046	92.6	90.6	97.9
25-44	1,045,017	980,657	93.8	92.7	98.8
45-64	1,257,580	1,197,505	95.2	94.9	99.7
65-74	376,531	359,589	95.5	95.3	99.8
75+	309,835	296,567	95.7	95.6	99.9
Sex					
Male	2,091,391	1,971,169	94.3	92.1	97.7
Female	2,234,645	2,112,516	94.5	92.7	98.0
Mode					
Mail	2,972,654	2,922,566	98.3	96.7	98.4
CATI	545,695	455,290	83.4	81.3	97.5
CAPI	807,687	705,829	87.4	84.0	96.1
CAPI Direct	74,570	63,900	85.7	83.7	97.7
CAPI Nonresponse	733,117	641,929	87.6	84.0	95.9

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Relative to younger age groups, older age groups (45 and over) had higher proportions of person records that were PIKed as well as higher match ratios. The difference was more sizable when it came to match ratios. Those 65 and above in the ACS were assigned PIKs at a rate of almost 96 percent and virtually all of those with PIKs (99.8 percent) were found in administrative records.[37] Administrative record coverage of the 45 to 64 age group with PIKs was also nearly exhaustive at 99.7 percent.

The 18 to 24 age group in the ACS had the lowest PIK rate of all age groups at 92.6 percent and the third lowest percentage of person records in administrative records (90.6 percent). However, administrative record coverage of those aged 18 to 24 with a PIK was considerably higher at almost 98 percent.

CATI and CAPI respondents in the 2010 ACS had lower PIK rates (83.4 and 87. 4 percent respectively) than mail respondents (98.3 percent), and the PIK rate among the two CAPI subcategories was very similar.[38] Almost 17 percent of CATI and 13 percent of CAPI respondents were not assigned a PIK. This result is consistent with those in the previous section, which suggested that relative to mail respondents, name and date of birth data collected from CATI/CAPI householders may be less reliable or absent, affecting PIK assignment. Not surprisingly, match ratios were lower for CATI/CAPI respondents than for mail respondents – 96.7 percent compared to 81.3 percent and 84.0 percent for CATI and CAPI respondents respectively. Again, the match rates of CAPI Direct and CAPI Nonresponse respondents were virtually the same. When only those with PIKs were taken into consideration, the match ratios across response modes (including CAPI subcategories) became close in magnitude.

Administrative records covered a substantial proportion of the Hispanics surveyed in the 2010 ACS. Still, persons of Hispanic origin in the 2010 ACS obtained PIKs and were found in administrative records at lower rates relative to non-Hispanics. The PIK assignment rates for Hispanics and non-Hispanics were 89.3 percent and 95.1 percent respectively. Meanwhile, the match ratio was 84.3 percent for Hispanics and 93.6 percent for non-Hispanics. When only Hispanic persons with PIKs were taken into consideration, their match ratio increased by 10 percentage points and the difference in match ratios by Hispanic origin was reduced to 3.8 percentage points.

Across race groups,[39] the percentage of persons PIKed in the 2010 ACS ranged from 86.9 percent for the SOR alone group to 95.0 percent for the White alone group. The match ratio ranking was the same as the ranking of percentage with PIKs - although the levels were lower. The SOR alone group had the lowest percentage of records PIKed at 86.9 percent and also had the lowest match ratio at 81.4 percent. This lower PIK assignment percentage for the SOR alone group was largely driven by the Hispanic group, as nearly 96 percent of those classified as SOR alone in the 2010 ACS were of Hispanic origin.[40]
No differences were found by sex in the percentage of PIKed person records or in the percentage of those found in both the 2010 ACS and administrative records.

4.5. Region

Table 9 shows the percentage of 2010 ACS person records with PIKs, and the 2010 ACS and administrative records person match ratios by region. The match ratios are expressed both as a percentage of all persons in the 2010 ACS and as a percentage of only those with PIKs. Note that the source of the region variable in Table 9 is the 2010 ACS.

---

[37]The extensive administrative records coverage of those aged 65 and above can be largely attributed to the inclusion of Medicare Part B in AR.
[38] Match results by mode subcategories across demographic characteristics are presented and discussed in Appendix V.
[39] Race categories were created following the methodology employed in the 2010 CMS: "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, AIAN alone, Asian alone, NHPI alone, and SOR 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." See page 6 of the 2010 CMS.
[40] This was also the case in the 2010 CMS.

The percentages of 2010 ACS person records PIKed did not vary a lot by region. They ranged from 92.9 percent in the West to 95.9 percent in the Midwest (a 3 percentage point difference). The match ratios, expressed as a percentage of all 2010 ACS person records, followed the same pattern. They ranged from 90.3 percent in the West to 94.0 percent in the Midwest. The Northeast had the second highest match ratio with 92.7 percent followed by the South with 92.2 percent.

This regional ranking of match ratios is different from the one observed in addresses with the exception of the Midwest – which was at the top for both addresses and persons. In contrast to the person results, the South and West held higher address match ratios than the Northeast.

It is worth noting that when the match rate was expressed as a percentage of 2010 ACS person records with PIKs, the variation in match ratios by region was reduced to just 0.9 percentage points, and the levels became considerably higher (in the high 90s) for all regions. Furthermore, the match ratio of the Midwest was virtually undistinguishable from those of the Northeast and the South. This suggests that some of the regional differences observed in match ratios (when expressed as a percentage of all 2010 ACS person records) were due to differences in PIK assignment rates, and not necessarily to administrative records undercoverage per se.

Table 9. 2010 ACS and Administrative Records Person Match by Region
Universe: 2010 ACS persons living in housing units

ACS Region	2010 ACS Persons	2010 ACS Persons with a PIK		Match of 2010 ACS and Administrative Records		
	Count	Count	As Percent of 2010 ACS Persons	Count	Match Ratio as % of ACS Person Count	Match Ratio as % of ACS PIK Count
Total	4,326,036	4,083,685	94.4	3,997,069	92.4	97.9
Northeast	808,598	763,215	94.4	749,200	92.7	98.2
Midwest	1,162,122	1,114,060	95.9	1,092,811	94.0	98.1
South	1,433,145	1,349,763	94.2	1,321,925	92.2	97.9
West	922,171	856,647	92.9	833,133	90.3	97.3

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data.

4.6. State by Federal and Commercial Administrative Records Data

Table 10 shows the percentage of persons in the ACS with PIKs, and the 2010 ACS and administrative records person match ratios by state. It also highlights the contribution of commercial data to administrative records coverage. Specifically, it presents state-level match ratios for federal administrative records (fourth column) as well as for combined federal and commercial administrative records (fifth column). Finally, the last column shows the increase in match ratios attributable to commercial data.

Match ratios (column 5) ranged from 87.3 percent in New Mexico to 95.4 percent in Iowa - an 8.0 percentage point difference. Consistent with the results by region, the five states with the highest match ratios were from the Midwest: Iowa, Minnesota, North Dakota, Ohio and Wisconsin.[41] Out of the five states with the lowest match ratios, four were Western states (New Mexico, Alaska, California and Hawaii) and the fifth one was the District of

[41] Pennsylvania, Nebraska and Kentucky had the same match ratio as Nebraska at 94.6 percent.

Columbia.[42] These five states had a higher percentage of CAPI Nonresponse and Hispanic respondents relative to the rest of the states.[43] When taking into account only person records with PIKs, the match ratio levels increased to the high 90s for all states and the spread across them decreased to just 2.7 percentage points (data not shown). As mentioned in the previous section, this suggests that a considerable portion of the observed geographic variation can be attributed to differences in PIK assignment rates across geography.

The incorporation of commercial administrative records increased the match ratios of all states, but rather modestly. On average, the increase was approximately 1 percentage point. Across states, increases ranged from a negligible 0.1 percentage point in Alaska to 1.5 percentage points in Michigan. Out of the 19 states experiencing increases of 1 percentage point or more, 15 of them were in the South and had match ratios in the bottom half of the distribution.

Relative to 2010 ACS persons found in federal administrative records, those matching to only commercial data had higher percentages of CAPI respondents, persons with ages between 25 and 64, the Black alone group, and persons living in the South (see Appendix II, Table 2).

The increases in the overall and state-level match ratios attributable to commercial administrative records were substantially higher for addresses than for persons. For addresses, they ranged from 5 percentage points in Maryland to 16.4 percentage points in Wyoming. This suggests that commercial administrative records data used in this study provides a sizable addition to address coverage of Federal sources, but that its marginal contribution to person coverage is not as relevant.

[42] Arizona, also a Western state, virtually tied the District of Columbia with a match ratio of 90.3 percent.
[43] Specifically, in the 2010 ACS response data sample, these five states had 22.2 percentage points more Hispanics and 6.5 percentage points more CAPI Nonresponse respondents than the rest of the states. Please remember these numbers are unweighted.

Table 10. 2010 ACS and Administrative Records Person Match by State
Universe: 2010 ACS persons living in housing units

ACS State	2010 ACS Person Records	Percent of 2010 ACS Person Records with PIKs	Match Ratio of 2010 ACS and Federal Administrative Records*	Match Ratio of 2010 ACS and Federal/ Commercial Administrative Records*	Percentage Point Increase in Match Ratio attributable to Commercial Data
Total	4,326,036	94.4	91.5	92.4	0.9
Alabama	67,333	94.6	92.0	93.1	1.2
Alaska	12,085	89.9	88.8	89.0	0.1
Arizona	72,466	92.7	89.4	90.3	0.9
Arkansas	40,984	94.6	91.9	92.9	1.1
California	455,196	92.3	88.3	89.3	1.0
Colorado	67,983	94.0	91.1	91.9	0.8
Connecticut	47,684	94.3	92.0	92.6	0.7
Delaware	12,989	95.9	93.2	94.2	1.0
District of Columbia	6,970	92.5	89.2	90.2	1.1
Florida	206,307	93.6	90.5	91.7	1.3
Georgia	113,287	93.2	89.9	91.1	1.2
Hawaii	19,025	91.8	89.3	89.9	0.7
Idaho	23,564	93.5	90.5	91.2	0.8
Illinois	188,271	95.0	92.5	93.1	0.7
Indiana	95,882	95.8	93.5	94.2	0.6
Iowa	62,860	96.7	94.8	95.4	0.5
Kansas	49,010	95.8	93.5	94.0	0.5
Kentucky	61,880	96.1	93.5	94.6	1.1
Louisiana	59,907	93.6	90.8	92.0	1.2
Maine	26,520	95.2	92.9	93.8	0.9
Maryland	72,866	95.3	92.5	93.3	0.8
Massachusetts	85,032	95.2	92.8	93.5	0.6
Michigan	176,294	96.1	91.0	92.5	1.5
Minnesota	125,672	96.5	94.7	95.2	0.5
Mississippi	37,344	95.3	92.3	93.6	1.3
Missouri	94,291	95.4	93.0	93.9	0.9
Montana	18,119	92.8	90.7	91.3	0.6
Nebraska	36,858	96.0	94.2	94.6	0.4
Nevada	29,996	93.7	90.1	91.1	1.0
New Hampshire	20,946	95.1	93.0	93.6	0.6
New Jersey	116,763	93.9	91.5	92.0	0.6
New Mexico	26,731	89.2	86.5	87.3	0.8
New York	265,509	92.8	90.0	90.8	0.8
North Carolina	120,135	94.2	91.2	92.3	1.1
North Dakota	16,177	95.9	94.6	94.9	0.3
Ohio	170,181	96.2	93.8	94.7	0.9
Oklahoma	62,149	94.1	91.4	92.4	1.0
Oregon	52,908	95.0	92.0	93.1	1.1
Pennsylvania	218,339	96.0	93.9	94.6	0.7
Rhode Island	12,934	95.0	92.8	93.5	0.7
South Carolina	56,305	94.3	91.5	92.6	1.1
South Dakota	17,070	95.1	93.5	93.7	0.2
Tennessee	81,952	95.1	92.4	93.5	1.1
Texas	304,569	93.5	89.9	90.9	1.1
Utah	42,244	94.1	90.7	91.2	0.5
Vermont	14,871	95.2	93.5	94.1	0.6
Virginia	101,395	94.9	92.3	93.1	0.8
Washington	93,593	94.7	92.1	92.9	0.8
West Virginia	26,773	95.1	92.5	93.7	1.2
Wisconsin	129,556	95.9	94.1	94.6	0.5
Wyoming	8,261	93.1	90.8	91.5	0.6

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data
* Match ratios are expressed as a percentage of all 2010 ACS person records.

4.7. 2010 Census Match Study

As discussed earlier in Section 2.9, although results from the 2010 ACS cannot be strictly compared to those from the 2010 CMS, a descriptive, non-statistical comparison is valuable to round off the overall assessment of administrative records coverage with respect to the ACS. With the strong caveats discussed in Section 2.9, this subsection discusses the person-level results in relation to those found in the 2010 CMS. Recall that the 2010 ACS data are unweighted and results reflect the presence or absence of administrative records for the persons in the survey response file. The 2010 ACS results are not intended to reflect the entire U.S. population.

The percentage of 2010 ACS person records with PIKs as well as match ratios in the 2010 ACS and the 2010 Census exhibited the same patterns across groups. The magnitudes were generally higher in the 2010 ACS with the exception of children 0 to 2 years old.[44,45] As already pointed out, this age group is a particularly hard to cover population with existing administrative records. This was also the case in the 2010 Census, but the problem was exacerbated by more pronounced time frame misalignments between the 2010 ACS sample and the administrative records at hand. The ACS is a rolling sample where households are interviewed throughout the year while administrative records used in this study (and the 2010 CMS) had reference dates of April 1, 2010 or prior. Thus, relative to the 2010 Census, the 2010 ACS had a higher proportion of children born in 2010, and in particular, after the Census reference date of April 1.

Not surprisingly, the lower coverage for the 0 to 2 age group was primarily driven by babies born in 2010 (i.e., age 0). Compared to children of ages 1 and 2 in the 2010 ACS, children that were born in 2010 had much lower match ratios – 49.9 percent versus 87.0 and 88.3 percent for children of ages 1 and 2 respectively (data not shown). Further examination also revealed that the proportion of ACS children born in 2010 who were also found in administrative records decreased as the month of interview was later in the year.

The person match ratios across regions exhibited the same pattern in the 2010 ACS and the 2010 Census. Regarding states, there was general agreement between the 2010 ACS and Census on the states with the lowest and highest person match ratios. Iowa, North Dakota, Wisconsin and Vermont were among the states with the highest match ratios, and New Mexico, Arizona, California, the District of Columbia and Nevada were among the states with the lowest match ratios in both the 2010 ACS and Census. Match ratios across states were larger in the 2010 ACS than in the 2010 Census.[46] The spread in the state-level match ratios was very similar in the two studies.[47]

5. Person-Address Pairs Count and Match Results

This section discusses administrative records coverage of persons at addresses in the 2010 ACS, and presents coverage results by type of living quarter, demographic characteristics, response mode and geographic area. It also assesses the contribution of commercial administrative records data to the coverage of person-address records in the 2010 ACS, and discusses results in relation to those found in the 2010 CMS. Match results by mode subcategories across demographic characteristics are presented and discussed in Appendix VI.

As in the address and person sections, the match numbers and ratios in the tables that follow represent the match of the 2010 ACS to both federal and commercial administrative records. However, in order to assess the contribution of commercial data to administrative records coverage, the state table presents state-level match ratios for federal and commercial data separately.

---

[44] Overall results held when both housing units and group quarters were included in the analysis.
[45] This age group had a match ratio 2.7 percentage points lower in the 2010 ACS.
[46] The only exception was Alaska, which had a slightly lower match ratio in the 2010 ACS (just 0.5 percentage points lower). This can be partly attributed to the higher proportion of CAPI respondents in that state relative to Census, and the associated lower PIK rate. The proportion of CAPI respondents in Alaska in the 2010 ACS was 51.3 percent – compared to 24.8 percent of NRFU in the 2010 Census. When only persons with PIKs were taken into account, Alaska's match ratio increased to 98.9 percent, becoming the second highest match ratio among states.
[47] A 10.7 percentage point difference in the 2010 Census versus an 8.0 percentage point difference in the 2010 ACS.

Also note that, as discussed in the Data and Methodology section, no best-address model was applied to administrative records data to try to identify the best address for a given person. Instead, all unique person-address pairs available in administrative records were used in the analysis.

As in the address and person coverage analysis sections, this section first presents general results for all living quarters (housing units and group quarters) followed by more detailed results for just housing units.

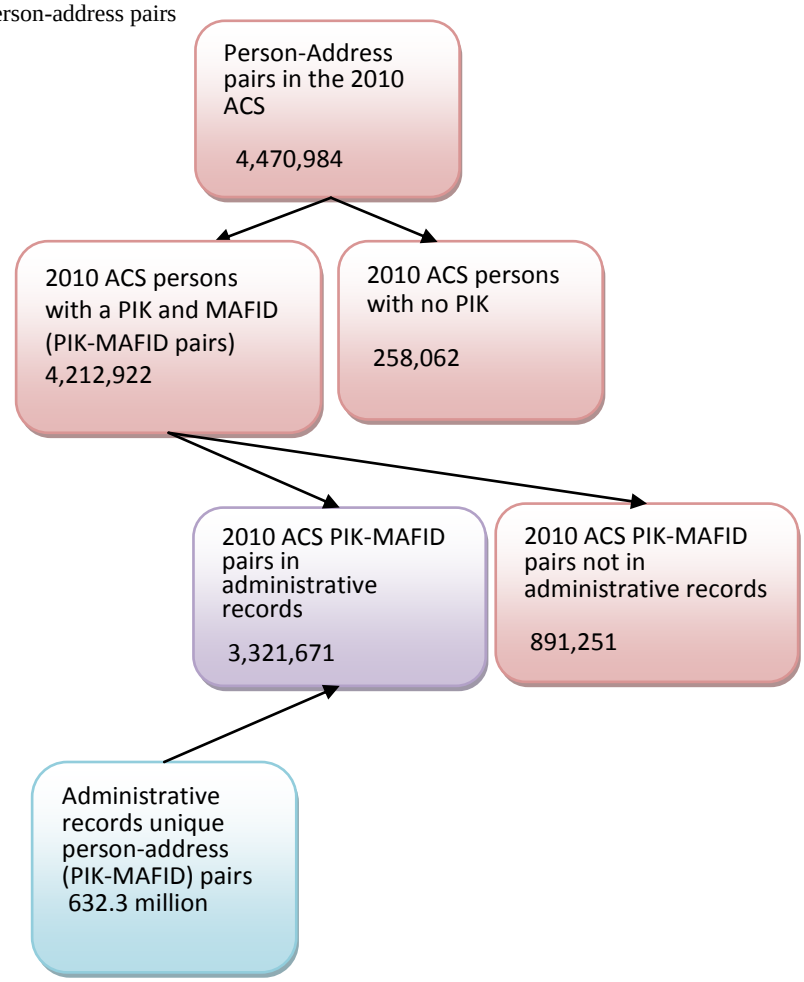
5.1. Overall

Figure 3 below shows the number and match of 2010 ACS and administrative records person-address pairs.[48] All persons in the 2010 ACS were associated with a MAFID, thus all 2010 ACS person and PIK counts discussed in the person section are the same as in Figure 2. For instance, there were 4.5 million people in the 2010 ACS and all of them had a MAFID associated with them.

There were 632.3 unique PIK-MAFID pairs in administrative records. Of the 4.47 million persons in the 2010 ACS (all with MAFIDs), 3.32 million (or 74.3 percent) matched to PIK-MAFID pairs in administrative records. These 3.32 million represented 78.8 percent of 2010 ACS persons with PIKs.

[48] As explained in the data section, the 2010 ACS included some 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 ACS person-address pairs.

Figure 3. Count and Match of 2010 ACS and Administrative Records Person-Address Pairs
Universe: All person-address pairs



Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data.

5.2. Housing Units and Group Quarters

Table 11 shows person-address record counts and match ratios for housing units and group quarters separately as well as overall. Compared to the overall address and person match ratios (92.7 and 92.1 percent respectively), the match ratio for person-address pairs was notably lower at 74.3 percent. Even when considering only 2010 ACS persons with PIKs, the match ratio remained relatively low at 78.8 percent.

A closer look at the match ratios of housing units and group quarters independently revealed a stark difference between them. The person-address match ratio for persons in housing units was 76.5 percent while that for persons in group quarters was just 7.1 percent. Even when considering only persons with PIKs, the difference in the match ratio between housing units and group quarters did not decrease – and in fact, became larger. These results suggest that the administrative records used in this study did not provide good coverage of addresses for people living in group quarters.

Table 11. 2010 ACS and Administrative Records Person-Address Count and Match by Type of Living Quarter

Universe: 2010 ACS person-address pairs

ACS Type of Living Quarter	2010 ACS person-address pairs*	2010 ACS person-address pairs with PIKs		Match of person-address pairs in 2010 ACS and Administrative Records		
	Count	Count	As a percent of person-address pairs	Count	Match Ratio as a percent of all ACS person-address pairs	Match Ratio as a percent of ACS person-address pairs with PIKs
Total	4,470,984	4,212,922	94.2	3,321,671	74.3	78.8
Housing Units	4,326,036	4,083,685	94.4	3,311,397	76.5	81.1
Group Quarters	144,948	129,237	89.2	10,274	7.1	7.9

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data.
* All persons in the 2010 ACS have MAFIDs.

While administrative records contained the vast majority of housing unit addresses and persons living in housing units in the 2010 ACS, they did not do as well at placing persons at the correct address. In particular, there were 0.7 million persons[49] in the 2010 ACS that were also found in administrative records for whom administrative records did not provide an address or the correct address. This represents approximately 18 percent of the 4.0 million persons living in housing units in the 2010 ACS that were also in administrative records.

Further examination revealed that persons in the 2010 ACS living in housing units for whom a person-address match was not found in administrative records tended to be CAPI respondents, more mobile,[50] and be under 18 years of age relative to those with a person-address match in administrative records.[51]

There are additional factors that contributed to the relatively low person-address match ratio. One regards time frame misalignment issues where the day the ACS household is interviewed fell outside the time period covered by the administrative records used in this study. As already mentioned, the administrative records used in this study have reference dates as of April 1, 2010 or prior while 2010 ACS data are collected throughout the 2010 calendar year. An additional contributing factor is ACS' residency rules. According to the ACS, anyone who is currently living or staying at the sample address for more than two months is considered a resident of that address. Therefore, the ACS may place persons at addresses that are not their usual residence (as long as they stay there for more than two months). By contrast, administrative records generally include a person's usual place of residence.

As in the address and person sections, the remainder of this section presents results only for housing units.

[49] The 0.7 million is derived from the 4.0 million persons in housing units that were in both the 2010 ACS and administrative records minus the 3.3 million persons in the 2010 ACS (living in housing units) that were found with the same address in administrative records.

[50] Mobility is measured by the percentage of persons over 1 year of age that did not live in the same house or apartment a year prior to answering the survey.

[51] Of those 2010 ACS persons living in housing units without a person-address match in administrative records, 31.1 percent of them had moved in the last year, 38.3 percent came from CAPI respondent housing units and 48.9 percent were under 18 years old. By contrast, only 7 percent of those with a person-address match had moved the prior year, 12.4 percent came from CAPI respondent housing units and 26.3 percent were under 18.

5.3. Demographic Characteristics and ACS Mode

Table 12 shows 2010 ACS and administrative records person-address match ratios by ACS race, Hispanic origin, age, sex and mode subcategories. As mentioned earlier, all persons in the 2010 ACS are associated with an address (i.e., MAFID), thus the count and percentages of 2010 ACS persons with PIKs and MAFIDs are the same as in the person analysis, and therefore, are not reproduced in this table.

Compared to the person results, the match ratios for person-address records were considerably lower across all demographic and mode groups. In addition, the spread of the person-address match ratios was larger than in the person results – more substantially for the race, age and mode groups than for the Hispanic origin and sex categories.

The groups with the lowest person-address match ratios were CAPI respondents, children ages 0 to 2 and the AIAN group. This was partly due to these three groups having the lowest (or among the lowest) address and/or person match ratios.

As in the address and person results, a higher percentage of 2010 ACS person-address pairs in mail matched to administrative records (84.8 percent) compared to CATI (69.9 percent) and CAPI (50.8 percent). A closer look at the CAPI subcategories reveals that the CAPI Direct group has a much lower match rate than the CAPI Nonresponse (although this latter subcategory still has a match ratio that is substantially lower than CATI and mail). The difference in match ratios between mail and CATI/CAPI was considerably higher for person-address pairs than for address and person records.

The difference in the match ratio between mail and CATI was largely due to the lower PIK assignment rate of CATI respondents. When the match ratio was calculated as a percentage of those person-address pairs with PIKs, the match ratio for CATI respondents increased to 83.8 percent, which was just 2.4 percentage points lower than the match ratio of mail respondents. However, this was not the case for CAPI respondents overall and CAPI subcategories. Their match ratios, expressed as a percentage of those person-address records with PIKs, remained among the lowest of any group in Table 12. In particular, the match ratio for the CAPI Direct category was the lowest of any group by far. This pattern was also observed in, and also partly driven by, address match ratios. The match ratio for CAPI Direct addresses was approximately 68 percentage points lower than the one for mail and CATI housing units.

Further examination revealed that relative to mail respondents, CAPI Nonresponse respondents tended to have a higher share of Hispanics, Black alone, SOR alone (most of whom are of Hispanic origin), younger persons and movers. Relative to the mail category, the CAPI Direct group had a higher share of AIAN. This is to be expected since the CAPI Direct group contains all households in Remote Alaska in the 2010 ACS response sample (see Appendix II, Table 1).

Table 12. 2010 ACS and Administrative Records Person-Address Match by Demographic Characteristics and Mode

Universe: 2010 ACS person-address pairs (only housing units)

ACS Demographic Characteristics and Mode	2010 ACS Person-Address Records	Match Count and Ratios of 2010 ACS and Administrative Records Person-Address Records		
	Count	Count	As Percent of all ACS person-address pairs	As Percent of ACS person-address pairs with PIKs
Total	4,326,036	3,311,397	76.5	81.1
Hispanic or Latino Origin				
Hispanic	542,890	350,415	64.5	72.3
Non Hispanic	3,783,146	2,960,982	78.3	82.3
Race				
White Alone	3,418,850	2,688,923	78.6	82.8
Black Alone	412,411	287,389	69.7	74.9
American Indian or Alaska Native Alone	41,886	21,520	51.4	56.8
Asian Alone	189,241	144,000	76.1	81.9
Native Hawaiian or Other Pacific Islander Alone	5,919	3,789	64.0	71.3
Some Other Race Alone	146,981	89,314	60.8	69.9
Two or More Races	110,748	76,462	69.0	73.2
Age				
0-2	152,665	81,992	53.7	57.4
3-17	848,408	571,587	67.4	71.8
18-24	336,000	218,141	64.9	70.1
25-44	1,045,017	798,466	76.4	81.4
45-64	1,257,580	1,054,608	83.9	88.1
65-74	376,531	321,854	85.5	89.5
75+	309,835	264,749	85.4	89.3
Sex				
Male	2,091,391	1,588,184	75.9	80.6
Female	2,234,645	1,723,213	77.1	81.6
Mode				
Mail	2,972,654	2,519,469	84.8	86.2
CATI	545,695	381,473	69.9	83.8
CAPI	807,687	410,455	50.8	58.2
CAPI Direct	74,570	7,317	9.8	11.5
CAPI Nonresponse	733,117	403,138	55.0	62.8

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data.

Across age groups, person-address match ratios ranged from 53.7 percent to 85.5 percent. These results followed the same general pattern as the person results for age, where match ratios were higher for the older age groups and lower for younger ones. The age groups 65 to 74, and 75 and over had the highest, and almost identical match ratios (85.5 and 85.4 percent respectively), followed by those aged 45 to 64 (83.9percent). The age group 0 to 2 had the lowest match ratio at 53.7percent followed at a distance by the 18 to 24 age group at 64.9 percent. Just as in the person analysis, the low person-address match ratio for the 0 to 2 age group was, to a great extent, driven by administrative records under-coverage of the youngest group and time frame misalignments that primarily affected those born in 2010 (i.e., age 0). Specifically, the person-address match ratio for babies born in 2010 was just 35.7 percent compared to 61.0 percent for those of age 1 and 63.0 percent for those age 2 (data not shown).

A higher percentage of non-Hispanic person-address records in the 2010 ACS (78.3 percent) matched to administrative records relative to Hispanics (64.5 percent). Likewise, the White alone group also had the highest person-address match ratio with 78.6 percent. The ranking of match ratios by race group changed relative to the person results though. The Asian alone, Native Hawaiian or Other Pacific Islander alone (NHPI) and SOR alone groups had higher rankings relative to the person results. In contrast, the Black alone, Two or more races, and particularly, the AIAN alone groups had lower rankings when compared to their person results.

5.4. Region

Table 13 presents the 2010 ACS and administrative records person-address match by region. The second column displays the 2010 ACS person-address count, the third column shows the number of matched 2010 ACS and administrative records person-address pairs, and the last two columns display the two versions of match ratios. As stated earlier, because all person records in the 2010 ACS have a MAFID, the number and percentage of person-address records with PIKs by region are the same as in the person analysis section, and therefore, are not reproduced in this table.

Table 13. 2010 ACS and Administrative Records Person-Address Match by Region
Universe: 2010 ACS person-address pairs (only housing units)

ACS Region	2010 ACS Person-Address Records	Match of 2010 ACS and Administrative Person-Address Records	Match Ratio of 2010 ACS and Administrative Person-Address Records	Match Ratio of 2010 ACS and Administrative Person-Address Records with PIKs
	Count	Count	As percent of all ACS person-address records	As percent of ACS person-address records with PIKs
Total	4,326,036	3,311,397	76.5	81.1
Northeast	808,598	620,838	76.8	81.3
Midwest	1,162,122	933,783	80.4	83.8
South	1,433,145	1,077,892	75.2	79.9
West	922,171	678,884	73.6	79.2

Sources: Unweighted 2010 ACS and 2010 Administrative Records Data.

The ranking of the match ratio by region follows that of the person analysis. The Midwest had the highest match ratio (80.4 percent) followed by the Northeast (76.8 percent), the South (75.2 percent) and finally the West (73.6 percent). The spread by region (6.8 percentage points) was higher than the one found for persons (3.0 percentage points) and for addresses (4.0 percentage points). This was largely due to the Midwest having a person-address match ratio almost 4 percentage points above the second highest.

When considering the match ratio as a percentage of 2010 ACS person-address records with PIKs, the spread by region was reduced by 2 percentage points with the levels for the South, West and Northeast becoming closer in magnitude. Just as with the person results, this seems to indicate that some of the differences observed across regions in person-address matches are due to differences in PIK assignment rates.

5.5. State by Federal and Commercial Data

Table 14 shows the 2010 ACS and administrative records person-address match by state, and also highlights the contribution of commercial data to state-level match ratios. The second column displays the 2010 ACS person-address count, the third column shows the person-address match ratio of 2010 ACS and Federal administrative records, the fourth displays the person-address match ratio of 2010 ACS and combined federal and commercial administrative records, and the last column presents the percentage point increase in the person-address match ratios that can be attributed to commercial administrative records.

The match ratios of combined federal and commercial administrative records and the 2010 ACS (column 3) ranged from 45.6 percent in Alaska to 83.2 percent in Ohio - a 37.6 percentage point difference. Four of the five states with the lowest match ratios were in the West and one was in the South: Alaska (45.6 percent), New Mexico (58.9 percent), West Virginia (61.4 percent), Montana (61.4 percent) and Wyoming (61.9 percent). These five states were among the ten states with the lowest address match ratios, and two of them (Alaska and New Mexico) had the lowest person match ratios as well. In particular, in the 2010 ACS response data sample, these five states had a higher percentage of CAPI Direct and AIAN person-address pairs than the rest of the states.[52]

Alaska's match ratio was substantially lower than the next state, New Mexico, which had a match ratio 13.3 percentage points higher. Compared to the other four states in this group, Alaska had a considerably higher percentage of CAPI Direct respondents,[53] and along Wyoming, was among the top five states with the highest rates of mobility in the U.S.[54] (data not shown).

[52] Specifically, in the 2010 ACS response data sample, these five states had 12.7 percentage points more CAPI Direct and 7.3 percentage points more AIAN respondents than the rest of the states. Please remember these numbers are unweighted.
[53] As mentioned earlier, this is expected since all Remote Alaska addresses in the 2010 ACS are subsampled directly into CAPI, and in addition, Alaska has a high percentage of unmailable/undeliverable addresses.
[54] Caution should be exercised if extrapolating 2010 mobility patterns into the future. In 2010, within and across state mobility flows were still influenced by the Great Recession. In turn, these mobility patterns are likely to have affected state-level person-address match ratios. Future research using data from non-recessionary time periods may reveal different mobility and person-place match ratio patterns.

Table 14. 2010 ACS and Administrative Records Person-Address Match by State
Universe: 2010 ACS person-address pairs (only housing units)

ACS State	2010 ACS Person-Address Records	Match Ratio of 2010 ACS and Federal Administrative Person-Address Records	Match Ratio of 2010 ACS and Combined Federal and Commercial Administrative Person-Address Records	Percentage Point Increase in Match Ratio attributable to Commercial Data
Total	4,326,036	76.1	76.5	0.5
Alabama	67,333	74.9	75.4	0.6
Alaska	12,085	45.6	45.6	0.0
Arizona	72,466	69.1	69.5	0.4
Arkansas	40,984	70.5	71.0	0.5
California	455,196	74.9	75.5	0.5
Colorado	67,983	75.7	76.1	0.4
Connecticut	47,684	80.7	81.1	0.4
Delaware	12,989	80.4	81.0	0.6
District of Columbia	6,970	73.3	73.9	0.6
Florida	206,307	77.4	78.2	0.8
Georgia	113,287	73.1	73.7	0.5
Hawaii	19,025	65.2	65.4	0.2
Idaho	23,564	69.6	69.9	0.3
Illinois	188,271	77.4	77.7	0.3
Indiana	95,882	79.8	80.2	0.3
Iowa	62,860	81.7	82.0	0.3
Kansas	49,010	77.9	78.2	0.3
Kentucky	61,880	76.7	77.2	0.6
Louisiana	59,907	73.1	73.7	0.6
Maine	26,520	71.5	71.9	0.4
Maryland	72,866	82.4	82.9	0.5
Massachusetts	85,032	80.1	80.5	0.4
Michigan	176,294	80.1	81.0	1.0
Minnesota	125,672	81.9	82.2	0.3
Mississippi	37,344	68.7	69.3	0.6
Missouri	94,291	77.3	77.7	0.4
Montana	18,119	61.2	61.4	0.2
Nebraska	36,858	76.4	76.6	0.2
Nevada	29,996	73.0	73.5	0.4
New Hampshire	20,946	77.4	77.7	0.3
New Jersey	116,763	78.5	78.8	0.3
New Mexico	26,731	58.6	58.9	0.3
New York	265,509	71.1	71.5	0.4
North Carolina	120,135	75.7	76.2	0.5
North Dakota	16,177	72.9	73.0	0.1
Ohio	170,181	82.7	83.2	0.5
Oklahoma	62,149	66.6	67.0	0.4
Oregon	52,908	77.3	77.8	0.6
Pennsylvania	218,339	80.3	80.7	0.4
Rhode Island	12,934	78.2	78.5	0.3
South Carolina	56,305	76.5	77.1	0.5
South Dakota	17,070	72.1	72.1	0.1
Tennessee	81,952	78.0	78.6	0.6
Texas	304,569	71.9	72.5	0.5
Utah	42,244	74.1	74.4	0.3
Vermont	14,871	68.8	69.0	0.2
Virginia	101,395	79.7	80.1	0.4
Washington	93,593	76.8	77.2	0.4
West Virginia	26,773	61.0	61.4	0.4
Wisconsin	129,556	82.6	82.9	0.3
Wyoming	8,261	61.7	61.9	0.2

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data.

Four out of the five states with the highest person-address match ratios were in the Midwest and one was in the South: Ohio (83.2 percent), Wisconsin (82.9 percent), Maryland (82.9 percent), Minnesota (82.2 percent) and Iowa (82.0 percent). Taken as a group, these states tended to have a lower percentage of CAPI respondents, and AIAN and Hispanic groups compared to the five states with the lowest match ratios.[55] No notable differences in mobility rates were found across the top five and bottom five states (see Appendix III, Table 1). These five states were among the five states with either the highest person and/or highest address match ratios.[56]

The contribution of commercial administrative records data to person-address records coverage was negligible, ranging from 0.02 percentage points in Alaska to 1.0 percentage points in Michigan. Taken together with the commercial evaluation results in the address and person sections, this evidence suggests that commercial administrative records data used in this study provided considerable coverage for addresses that would not have been covered otherwise by Federal sources. However, commercial data did not seem to extend coverage for 2010 ACS persons or persons at addresses beyond what the Federal sources already provided.[57]

5.6. 2010 Census Match Study

With the strong caveats discussed in Section 2.9, this subsection discusses the person-address results in relation to those found in the 2010 CMS. Recall that the 2010 ACS data are unweighted and results reflect the presence or absence of administrative records for the persons at addresses in the survey response file. The 2010 ACS results are not intended to reflect all housing units or the entire U.S. address or person populations.

The overall person-address match ratio in the 2010 ACS was higher than in the 2010 Census.[58] This is partly attributable to the higher address-level and person-level match ratios in the 2010 ACS. In addition, this study makes use of all unique person-address pairs in administrative records while in the 2010 CMS, PIKs were assigned to addresses based on a best-address model.[59]

The 2010 CMS found similar match ratio patterns by demographic and mode groups. In general, match ratios across groups in the 2010 ACS were higher than in the 2010 Census with the exception of the 0 to 2 age category. For this group, the 2010 ACS match ratio was 2 percentage points lower than the one in the 2010 Census. This result was also observed in the person analysis, and was driven by the same time frame misalignment issues between the 2010 ACS sample and the administrative records at hand.[60]
Regarding region and states, the pattern of person-address match ratios by region in the 2010 ACS was the same as in the 2010 Census. The states with the five highest and lowest person-address match ratios followed a regional pattern similar to the one found in the CMS, with Midwestern states dominating the top five and Western states comprising the majority of the lowest five. Minnesota, Wisconsin, Iowa and Ohio were part of the top five states in both the CMS as well as this study. Likewise, Alaska, West Virginia, New Mexico and Wyoming were among the bottom five states in both studies. Again, state person-address match ratios were higher for the 2010 ACS with the exception of Alaska. The lower person-address match ratio for Alaska in the 2010 ACS can be largely explained by its lower address and person match ratios.

---

[55] The five states with the highest person-address match ratio had on average 0.5 percent of AIAN, 3.5 percent of Hispanics and 12 percent of CAPI respondents in the 2010 ACS response data. By contrast, the five states with the lowest person-address match ratio had on average 8.2 percent of AIAN, 14.2 percent of Hispanics and 32.2 percent of CAPI respondents in the 2010 ACS response data.
[56] In particular, Ohio and Iowa were among the states with the highest person and address match ratios, Maryland had the highest address match ratio and Minnesota had the second highest person match ratio.
[57] Research will assess the coverage and quality of commercial data for demographics characteristics in the 2010 ACS. In addition, research currently underway is examining the contribution of commercial data overall and for specific groups.
[58] Specifically, 74.3 percent versus 65.8 percent.
[59] Nevertheless, when the person-address match ratio was calculated using the best-address model employed in the 2010 CMS, the person-address ratio in the 2010 ACS remained relatively high at 68.0 percent.
[60] On the other hand, the match ratio of the 18 to 24 age group was considerably higher in the 2010 ACS. This could be related to the fact that this analysis did not include group quarters and did not attempt to select a best address for a given person. The 18 to 24 age group includes populations, such as students, who are more likely to live in group quarters, and is also generally more mobile than other age groups.

6. Conclusions

Overall, administrative records provided substantial coverage for addresses and persons in the 2010 ACS (92.7 and 92.1 percent respectively), and less extensive though substantial coverage for person-address pairs (74.3 percent). When only person records with PIKs were taken into consideration, the overall match ratio for persons increased to nearly 98 percent, and for person-address records to almost 79 percent.

The 2010 CMS also found higher administrative records coverage for address and persons than for person-address pairs. Meanwhile, administrative records coverage of the ACS data was higher than in the 2010 Census overall and for most demographic, geographic and response mode groups.[61]

Administrative records also covered a significant proportion of the 2010 ACS CATI and CAPI addresses and respondents. Approximately 97 percent of CATI addresses and 85.4 percent of CAPI addresses in the 2010 ACS were found in administrative records. As expected, though, the address match ratio for CAPI Direct addresses was a lot lower since these are unmailable/undeliverable or Remote Alaska addresses and those are either hard to match to or not present in administrative records data. The person match ratio was 81.3 percent for persons living in CATI households and 84.0 percent for CAPI respondents. When only person with PIKs were considered, the person match ratios increased to 97.5 and 96.1 percent for CATI and CAPI respectively. When it came to person analysis, the two CAPI subcategories did not exhibit large differences.

On the other hand, the address, person and/or person-address coverage of certain groups in the 2010 ACS response data was lower in administrative records; in particular, young children, AIAN, SOR, NHPI, Hispanics as well as group quarters. Extending administrative records address and/or person coverage for these groups in the ACS response data requires the acquisition of additional administrative records data sources with good coverage of these groups.[62]

On this front, commercial data appeared to make a contribution to overall address coverage, but their contribution to person and person-address coverage was negligible. Concurrent research on commercial data quality is underway.[63] In addition, acquiring and making use of multiple years of administrative records data could help improve person-address coverage since adding a longitudinal dimension to administrative records would improve coverage of people with higher mobility. In particular, a closer examination of the CAPI Nonresponse subcategory (who had the lowest person-address match ratio along with the CAPI Direct group) revealed that they had a higher proportion of movers compared to their mail or CATI counterparts.

Results also showed a stark coverage disparity between housing units and group quarters for addresses, but particularly, for person-address records. Only 7.9 percent of person-address records for those living in group quarters were matched to administrative records, compared to 81.1 percent for those living in housing units. Administrative records coverage of housing unit addresses in the 2010 ACS was 93.3 percent relative to 54.4 percent for group quarters. In order to improve administrative records address coverage of group quarters, other sources of administrative records data need to be acquired.

The lower PIK assignment rates of certain groups[64] played a considerable role in the lower person match ratios as well as in the observed differences across demographic groups, response modes, regions and states. Hence, understanding how to reduce differences in PIK assignment rates as well as exploring new record linkage methodologies not dependent on PIK assignment would translate into better administrative records coverage of

[61] This could be partly due to the fact that the share of non-mail records is higher in the 2010 Census than in the 2010 ACS, and that non-mail records are less likely to be present in administrative records data relative to mail.
[62] This recommendation was also stated in the 2010 CMS. Also, as stated earlier, changes to the ACS sample as the decade progresses will be taken into consideration when evaluating administrative records coverage of the ACS.
[63] The 2010 CMS found that commercial data quality on race and Hispanic origin was inferior to federal data.
[64] These included CATI and CAPI respondents, the SOR group, and to a lesser extent, the NHPI, AIAN and Hispanic groups.

persons in the 2010 ACS.[65] Exploiting the wealth of information provided by the ACS, further research on unvalidated person records aimed to improve PIK assignment rates is to be undertaken.[66]

Persons living in CATI and CAPI households had the lowest PIK assignment rates of any other demographic or mode group. This suggests that, relative to mail respondents, name and date of birth data collected from CATI/CAPI householders may be less reliable or absent, affecting PIK assignment. Efforts to collect better personal identifiable information from non-mail households in the short-term would likely improve PIK assignment rates and match ratios, which in turn, would facilitate the evaluation of administrative records coverage of CAPI/CATI households.

Findings also indicated that low address match ratios were in part driven by vacant housing units, and that the overwhelming majority of vacant housing units (85 percent) were CAPI addresses. Research currently underway is assessing administrative records ability to identify vacant units as well as deletes. Using administrative records to resolve the vacant or delete status of a housing unit would improve the address frame of the ACS as well as realize considerable costs savings.

Findings from this initial study indicated that administrative records provide substantial coverage of 2010 ACS addresses, persons and person-address pairs. Such administrative records coverage of the 2010 ACS is a first step in supporting the use of the ACS in future administrative records coverage research. Among other things, upcoming research will assess administrative records coverage of the underlying population(s) that the ACS is designed to represent and the characteristics associated with the likelihood of such coverage.[67] Administrative records research will also include the evaluation of the quality and coverage of Hispanic origin, race, sex, and age data in administrative records relative to the 2010 ACS, and the assessment of administrative records ability to assist ACS address frame as well as ACS CATI and CAPI operations.

[65] These modes of action were also recommended in the 2010 CMS.
[66] To this end, further exploration of the CAPI and CATI groups is warranted as they comprised nearly 80 percent of person records without PIKs.
[67] This type of research will employ multivariate regression analysis and ACS weights.

References

Alvey, Wendy and Fritz Scheuren. 1982. "Background for an Administrative Record Census," Proceedings" Proceedings of the Social Statistics Section, American Statistical Association.

Asher, J. 2010. "A Cross-National Comparison of Population and Housing Census Methods." U.S. Census Bureau unpublished memorandum. December 22, 2010.

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.

Fallica, Heather, Sarah Heimel, Geoff Jackson, and Bei Zhang. 2012. "2010 Census Update Enumerate Operations Assessment Update Enumerate Production, Update Enumerate Quality Control, Remote Update Enumerate, and Remote Alaska," 2010 Census Program for Evaluations and Experiments, 2010 Planning Memoranda Series No. 245, U.S. Census Bureau.

Farber, James and Charlene Leggieri. 2002. "Building and Validating a National Administrative Records Database for the United States," New Zealand Conference on Database Integration.

Groen, Jeffrey. 2012. "Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures," Journal of Official Statistics, Vol.28, No.2, pp. 173–198.

Iams, Howard and Irena Dushi. 2009. "Pension Coverage and Pension Types: Using W-2 Tax Records to Correct SIPP Survey Reports." Paper presented at the 2009 Federal Committee on Statistical Methodology Research Conference. Washington, DC.

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.

Judson, D.H. and Bauder, Mark. 2003. "Administrative Records Experiment in 2000 (AREX 2000)", U.S. Census Bureau.

Meijer, Erik, Lynn A. Karoly, and Pierre-Carl Michaud. 2010. "Using Matched Survey and Administrative Data to Estimate Eligibility for the Medicare Part D Low Income Subsidy Program." Paper presented at the 2009 Federal Committee on Statistical Methodology Research Conference. Washington, DC.

Mulalic, Lada. 2011. "Dutch Virtual Census," Paper presented at the U.S. Census Bureau Conference on Utilizing Administrative Data: Technical, Statistical and Research Issues. Washington, D.C. October 27 & 28, 2011.

NORC 2011. "Assessment of the U.S. Census Bureau's Person Identification Validation System," Final Report. March 31, 2011.

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.

Rastogi, Sonya and Amy O'Hara. 2012. "Census Match Study," 2010 Census Program for Evaluations and Experiments, 2010 Planning Memoranda Series No. 247, 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.

U.S. Census Bureau. 2009. "Design and Methodology: American Community Survey." U.S. Government Printing Office, Washington, DC.

U.S. Census Bureau. 2008. "A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know." Washington DC: U.S. Government Printing Office.

U.S. Census Bureau. 2008. "2005 ACS Respondent Characteristics Evaluation." DSSD ACS Memo Series. http://www.census.gov/acs/www/Downloads/library/2008/2008_Joshipura_01.pdf.

U.S. Government Accountability Office. 2012. "Additional Actions Could Improve the Census Bureau's Ability to Control Costs for the 2020 Census." Report to Congressional Requesters. GAO-12-80.

Valente, Paolo. 2010. "Census taking in Europe: how are populations counted in 2010?" *Population and Societies*, No.467, May 2010.

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

Zanutto E., and A.M. Zaslavsky. Using administrative records to impute for nonresponse. In: Groves RM, Dillman DA, Eltinge JL, Little RJA, eds. *Survey Nonrespons*e. New York: John Wiley and Sons, 2002:403-15.

APPENDIX I

Appendix I, Table 1. 2010 ACS Address Match to Federal and Commercial Administrative Records by Urban/Rural Status and Housing Unit Type

ACS Characteristics	2010 ACS Addresses		2010 ACS Address that match to FEDERAL Administrative Records		2010 ACS Addresses that match to COMMERCIAL Administrative Records but do not match to Federal Records	
	Count	Percent	Count	Percent	Count	Percent
Total	1,917,799	100.0	1,638,811	100.0	144,894	100.0
Urban	1,273,065	66.4	1,157,252	70.6	72,845	50.3
Rural	644,734	33.6	481,559	29.4	72,049	49.7
Mobile home	125,795	6.6	89,900	5.5	14,339	9.9
Single family home	1,398,315	72.9	1,225,758	74.8	89,865	62.0
Building with 2-4 apartments	128,800	6.7	97,903	6.0	14,646	10.1
Building with 5 to 9 apartments	69,606	3.6	58,739	3.6	6,851	4.7
Building with 10 to 19 apartments	61,870	3.2	52,805	3.2	5,997	4.1
Building with 20+ apartments	131,946	6.9	112,883	6.9	12,946	8.9
Other (Boat/RV/van etc.)	1,467	0.1	823	0.1	250	0.2

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix I, Table 2. 2010 ACS Address Match to Federal and Commercial Administrative Records by Demographic Characteristics and Mode

ACS Characteristics	2010 ACS Addresses		2010 ACS Address that match to FEDERAL Administrative Records		2010 ACS Addresses that match to COMMERCIAL Administrative Records but do not match to Federal Records	
	Count	Percent	Count	Percent	Count	Percent
Total	1,750,859	100.0	1,567,243	100.0	99,093	100.0
Non Hispanic	1,597,219	91.2	1,431,227	91.3	89,117	89.9
Hispanic	153,640	8.8	136,016	8.7	9,976	10.1
White Alone	1,438,326	82.1	1,288,723	82.2	81,579	82.3
Black Alone	165,126	9.4	148,746	9.5	9,132	9.2
AIAN Alone	14,649	0.8	10,355	0.7	1,059	1.1
Asian Alone	62,106	3.5	57,396	3.7	2,558	2.6
NHPI Alone	1,677	0.1	1,487	0.1	105	0.1
SOR Alone	40,922	2.3	35,833	2.3	2,936	3.0
Two or more races	28,053	1.6	24,703	1.6	1,724	1.7
Mail	1,251,082	71.5	1,162,269	74.2	53,637	54.1
CATI	201,262	11.5	177,159	11.3	18,254	18.4
CAPI	298,515	17.0	227,815	14.5	27,202	27.5

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix I, Table 3. Percent of Housing Units in 2010 ACS under each Mode Category by State

	All Modes	MODE			CAPI SUBCATEGORIES	
ACS State	Count	% Mail	% CATI	% CAPI	% CAPI Direct	% CAPI Nonresponse
Total	1,917,799	65.8	11.2	23.0	3.8	21.7
Alabama	31,552	61.8	12.7	25.5	38.4	19.4
Alaska	5,604	35.6	6.5	57.9	6.4	24.6
Arizona	33,254	59.5	9.5	31.0	5.7	21.9
Arkansas	19,178	60.3	12.1	27.6	0.8	23.4
California	176,551	65.2	10.7	24.1	1.9	20.2
Colorado	30,885	67.0	10.9	22.1	0.3	15.9
Connecticut	20,503	70.4	13.4	16.2	2.3	22.0
Delaware	6,339	62.6	13.2	24.3	0.0	25.3
District of Columbia	3,819	65.4	9.2	25.3	0.6	25.9
Florida	100,820	63.0	10.5	26.5	2.8	23.9
Georgia	49,876	61.5	11.9	26.6	6.4	23.4
Hawaii	7,537	62.2	7.9	29.9	5.0	18.3
Idaho	10,126	64.2	12.5	23.3	1.0	19.2
Illinois	80,592	70.4	9.4	20.2	1.0	16.6
Indiana	42,080	71.3	11.2	17.5	0.6	12.4
Iowa	28,130	75.7	11.2	13.0	2.4	15.9
Kansas	22,065	69.7	12.0	18.2	4.4	16.6
Kentucky	28,208	67.2	11.8	21.0	2.9	24.0
Louisiana	27,108	57.7	15.4	26.9	20.9	17.3
Maine	15,212	49.8	12.1	38.1	0.4	16.9
Maryland	30,811	70.4	12.3	17.3	0.7	16.9
Massachusetts	37,494	70.3	12.0	17.6	0.0	25.3
Michigan	82,647	69.4	10.2	20.4	2.7	17.7
Minnesota	56,799	73.2	10.5	16.3	3.8	12.5
Mississippi	16,790	56.6	12.8	30.6	5.2	25.4
Missouri	43,440	68.1	10.9	21.0	4.6	16.4
Montana	9,256	53.0	14.0	33.0	14.0	19.1
Nebraska	16,768	71.3	11.1	17.7	4.0	13.7
Nevada	13,310	61.5	8.1	30.4	1.6	28.8
New Hampshire	10,013	62.4	12.5	25.2	7.7	17.5
New Jersey	48,420	68.0	11.6	20.5	0.2	20.2
New Mexico	12,779	50.2	10.9	38.9	15.4	23.5
New York	116,705	62.5	11.0	26.4	2.6	23.8
North Carolina	56,018	63.0	13.3	23.8	2.7	21.1
North Dakota	7,834	60.7	17.1	22.2	7.4	14.8
Ohio	76,268	71.4	10.8	17.8	0.6	17.2
Oklahoma	28,727	55.9	12.3	31.7	10.9	20.9
Oregon	23,620	70.6	10.7	18.7	1.4	17.4
Pennsylvania	100,139	70.1	10.3	19.6	5.6	13.9
Rhode Island	5,942	67.6	11.1	21.3	0.9	20.4
South Carolina	26,571	61.7	12.8	25.4	2.7	22.7
South Dakota	7,935	61.8	15.3	22.8	7.3	15.6
Tennessee	37,220	66.8	12.0	21.2	2.1	19.2
Texas	127,757	59.5	11.8	28.7	4.0	24.7
Utah	14,934	67.0	11.9	21.1	4.1	17.0
Vermont	7,833	57.0	12.8	30.1	10.8	19.3
Virginia	44,284	69.1	11.7	19.2	2.4	16.8
Washington	40,883	70.6	11.0	18.4	1.7	16.7
West Virginia	13,311	51.7	10.8	37.5	21.5	16.0
Wisconsin	59,938	74.3	10.1	15.7	1.0	14.7
Wyoming	3,914	54.0	17.0	29.0	6.1	22.9

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix I, Table 4. ACS 2010 and Administrative Records Address Count, Match Numbers and Ratios by Mode & Urban Status

ACS Mode by Urban Status	2010 ACS Address Count	2010 ACS and Administrative Records Address Match Count	2010 ACS and Administrative Records Address Match Ratio
Total	1,917,799	1,783,705	93.0
Urban	1,273,065	1,230,097	96.6
Rural	644,734	553,608	85.9
Mail	1,261,247	1,224,855	97.1
Urban	869,505	849,878	97.7
Rural	391,742	374,977	95.7
CATI	215,526	208,263	96.6
Urban	132,237	129,973	98.3
Rural	83,289	78,290	94.0
CAPI	441,026	350,587	79.5
Urban	271,323	250,246	92.2
Rural	169,703	100,341	59.1

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix I, Table 5. 2010 ACS and Administrative Records Address Count, Match Numbers and Ratios by Mode & Occupancy Status

ACS Mode and Occupancy Status	2010 ACS Address Count	2010 ACS and Administrative Records Address Match Count	2010 ACS and Administrative Records Address Match Ratio
Total	1,917,799	1,783,705	93.0
Occupied	1,750,859	1,666,336	95.2
Vacant	166,940	117,369	70.3
Mail	1,261,247	1,224,855	97.1
Occupied	1,251,082	1,215,906	97.2
Vacant	10,165	8,949	88.0
CATI	215,526	208,263	96.6
Occupied	201,262	195,413	97.1
Vacant	14,264	12,850	90.1
CAPI	441,026	350,587	79.5
Occupied	298,515	255,017	85.4
Vacant	142,511	95,570	67.1

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix I, Table 6. ACS 2010 Housing Units: Total, Occupied, Vacant, and Percent Vacant by State

ACS State	2010 ACS Address Count			
	All Housing Units	Occupied	Vacant	% Vacant
Total	1,917,799	1,750,859	166,940	8.7
Alabama	31,552	28,511	3,041	9.6
Alaska	5,604	4,462	1,142	20.4
Arizona	33,254	29,081	4,173	12.5
Arkansas	19,178	17,251	1,927	10.0
California	176,551	165,993	10,558	6.0
Colorado	30,885	28,444	2,441	7.9
Connecticut	20,503	19,419	1,084	5.3
Delaware	6,339	5,380	959	15.1
District of Columbia	3,819	3,471	348	9.1
Florida	100,820	88,041	12,779	12.7
Georgia	49,876	45,231	4,645	9.3
Hawaii	7,537	6,793	744	9.9
Idaho	10,126	9,157	969	9.6
Illinois	80,592	75,735	4,857	6.0
Indiana	42,080	39,373	2,707	6.4
Iowa	28,130	26,771	1,359	4.8
Kansas	22,065	20,562	1,503	6.8
Kentucky	28,208	26,116	2,092	7.4
Louisiana	27,108	24,598	2,510	9.3
Maine	15,212	11,523	3,689	24.3
Maryland	30,811	28,923	1,888	6.1
Massachusetts	37,494	35,047	2,447	6.5
Michigan	82,647	72,837	9,810	11.9
Minnesota	56,799	51,653	5,146	9.1
Mississippi	16,790	15,124	1,666	9.9
Missouri	43,440	39,750	3,690	8.5
Montana	9,256	7,864	1,392	15.0
Nebraska	16,768	15,600	1,168	7.0
Nevada	13,310	11,969	1,341	10.1
New Hampshire	10,013	8,665	1,348	13.5
New Jersey	48,420	45,099	3,321	6.9
New Mexico	12,779	10,969	1,810	14.2
New York	116,705	106,656	10,049	8.6
North Carolina	56,018	50,548	5,470	9.8
North Dakota	7,834	7,073	761	9.7
Ohio	76,268	71,282	4,986	6.5
Oklahoma	28,727	25,783	2,944	10.2
Oregon	23,620	22,131	1,489	6.3
Pennsylvania	100,139	91,434	8,705	8.7
Rhode Island	5,942	5,431	511	8.6
South Carolina	26,571	23,712	2,859	10.8
South Dakota	7,935	7,198	737	9.3
Tennessee	37,220	34,306	2,914	7.8
Texas	127,757	116,481	11,276	8.8
Utah	14,934	13,907	1,027	6.9
Vermont	7,833	6,425	1,408	18.0
Virginia	44,284	41,394	2,890	6.5
Washington	40,883	38,349	2,534	6.2
West Virginia	13,311	11,744	1,567	11.8
Wisconsin	59,938	54,108	5,830	9.7
Wyoming	3,914	3,485	429	11.0

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix I, Table 7. Match Ratios for All, Occupied, and Vacant Housing Units by State

ACS State	2010 ACS and Administrative Records Address Match Ratio		
	All Housing Units	Occupied	Vacant
Total	93.0	95.2	70.3
Alabama	92.7	94.6	74.3
Alaska	60.8	70.9	21.4
Arizona	89.7	92.7	68.6
Arkansas	90.5	92.9	69.0
California	96.6	97.6	81.8
Colorado	93.4	95.8	65.6
Connecticut	96.2	96.8	85.1
Delaware	93.2	95.7	79.1
District of Columbia	97.0	97.4	92.8
Florida	96.5	97.4	90.6
Georgia	93.7	95.1	80.5
Hawaii	87.1	89.6	64.9
Idaho	89.4	93.1	54.8
Illinois	95.1	95.9	82.5
Indiana	96.5	97.2	87.2
Iowa	97.0	97.5	85.2
Kansas	95.0	96.3	76.8
Kentucky	92.4	94.2	69.5
Louisiana	93.8	95.7	75.3
Maine	74.7	88.6	31.3
Maryland	97.4	98.0	87.7
Massachusetts	95.2	96.3	80.3
Michigan	93.8	97.3	67.5
Minnesota	92.4	96.0	56.4
Mississippi	90.2	92.1	73.0
Missouri	92.2	94.6	66.7
Montana	77.1	84.3	36.3
Nebraska	92.9	95.1	63.7
Nevada	94.7	96.0	83.7
New Hampshire	86.7	93.4	43.2
New Jersey	95.2	95.9	84.7
New Mexico	78.1	84.5	38.9
New York	88.7	91.3	61.1
North Carolina	93.3	95.2	75.0
North Dakota	86.3	90.6	46.0
Ohio	97.0	97.6	87.4
Oklahoma	86.1	89.2	58.5
Oregon	95.4	96.7	76.2
Pennsylvania	90.6	94.4	50.7
Rhode Island	94.0	95.5	78.3
South Carolina	93.6	95.7	76.8
South Dakota	87.5	91.2	52.1
Tennessee	95.3	96.6	80.4
Texas	92.7	94.6	73.0
Utah	91.8	94.6	53.3
Vermont	77.6	86.6	36.4
Virginia	95.3	96.7	75.2
Washington	95.0	96.6	72.0
West Virginia	75.5	80.1	41.2
Wisconsin	93.8	96.7	67.5
Wyoming	85.9	90.3	50.3

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

APPENDIX II

Appendix II, Table 1. Characteristics of Mail, CATI & CAPI Subcategory Respondents in the 2010 ACS

ACS Characteristics	MODE			CAPI SUBCATEGORIES	
	Mail	CATI	CAPI	CAPI Direct	CAPI Nonresponse
	Percent	Percent	Percent	Percent	Percent
Total	100.0	100.0	100.0	100.0	100.0
Hispanic Origin					
Hispanic	8.4	17.1	24.6	9.3	26.2
Non Hispanic	91.6	82.9	75.4	90.7	73.8
Race					
White Alone	83.0	77.4	65.6	75.2	64.6
Black Alone	7.2	11.9	16.6	6.0	17.7
AIAN Alone	0.6	0.9	2.2	12.7	1.2
Asian Alone	4.7	3.4	3.8	0.8	4.1
NHPI Alone	0.1	0.2	0.3	0.2	0.3
SOR Alone	2.0	3.5	8.6	2.0	9.3
Two or more races	2.4	2.7	2.9	3.1	2.9
Age					
0-2	3.2	3.1	5.1	4.0	5.2
3-17	17.8	22.	24.5	21.7	24.8
18-24	7.0	6.8	11.2	7.9	11.5
25-44	23.1	21.0	30.3	23.6	31.0
45-64	30.9	30.0	21.7	29.0	20.9
65-74	9.9	8.9	4.3	8.3	3.9
75+	8.2	7.9	2.9	5.5	2.6
Urban Status					
Urban	68.0	64.1	68.0	11.5	73.7
Rural	32.0	35.9	32.0	88.5	26.3
Movers					
Movers	10.8	6.2	19.2	10.6	20.0
Non-movers	89.2	93.8	80.8	89.4	80.0
Region					
Northeast	18.9	18.5	17.9	15.5	18.1
Midwest	29.2	24.9	19.6	15.7	20.0
South	31.6	35.4	37.0	44.6	36.3
West	20.2	21.2	25.5	24.2	25.6

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

Appendix II, Table 2. 2010 ACS Person Match to Federal and Commercial Administrative Records by ACS Demographic Characteristics, Mode, and Region

ACS Characteristics	2010 ACS Person Records	2010 ACS Person Records that match to FEDERAL Administrative Records	2010 ACS Persons Records that match to COMMERCIAL Administrative Records but do not match to Federal Records
	Percent	Percent	Percent
Total	100.0	100.0	100.0
Hispanic Origin			
Hispanic	12.5	11.4	14.3
Non Hispanic	87.5	88.6	85.7
Race			
White Alone	79.0	79.9	72.9
Black Alone	9.5	9.3	16.2
AIAN Alone	1.0	0.9	0.7
Asian Alone	4.4	4.3	2.9
NHPI Alone	0.1	0.1	0.1
SOR Alone	3.4	3.0	4.3
Two or more races	2.6	2.5	2.8
Age			
0-2	3.5	2.9	0.1
3-17	19.6	19.2	2.8
18-24	7.8	7.6	8.6
25-44	24.2	24.1	41.5
45-64	29.1	29.7	45.9
65-74	8.7	9.1	0.8
75+	7.2	7.5	0.3
Mode			
Mail	68.7	72.0	60.7
CATI	12.6	11.1	12.1
CAPI	18.7	16.9	27.2
Region			
Northeast	18.7	18.8	14.7
Midwest	26.9	27.4	23.3
South	33.1	33.0	40.6
West	21.3	20.8	21.3

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

APPENDIX III

Appendix III, Table 1. Characteristics of Persons in the States with the Lowest & Highest Person-Address Match Ratios

	Five States w/ Lowest Ratios	Five States w/ Highest Ratios
ACS Characteristics	Percent	Percent
Total	100.0	100.0
Hispanic Origin		
Hispanic	14.2	3.5
Non-Hispanic	85.8	96.5
Race		
White Alone	82.7	87.1
Black Alone	1.6	7.4
AIAN Alone	8.2	0.5
Asian Alone	1.3	2.2
NHPI Alone	0.1	0.02
SOR Alone	3.1	0.9
Two or more races	3.0	1.8
Age		
0-2	3.7	3.4
3-17	18.9	19.6
18-24	7.6	7.1
25-44	22.5	23.2
45-64	30.8	30.5
65-74	9.3	8.8
75+	7.1	7.3
Mode		
Mail	54.9	76.2
CATI	12.9	11.8
CAPI	32.2	12.0
Mobility Status		
Movers	11.5	10.2
Non- Movers	88.5	89.8
Urban/Rural Status		
Urban	51.0	55.0
Rural	49.0	45.0

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

APPENDIX IV

Appendix IV, Table 1. Address Match Ratios for Demographic groups by Mode & CAPI Subcategory, and Differences in Match Ratios from Mail Category

	MAIL		CATI		CAPI DIRECT		CAPI NONRESPONSE		Percentage Point Difference from Mail Match Ratio		
ACS Demographic Characteristics	Count	Match Ratio	Count	Match Ratio	Count	Match Ratio	Count	Match Ratio	CATI	CAPI Direct	CAPI Nonresp
Total	1,251,082	97.2	201,262	97.1	29,615	28.9	268,900	91.7	-0.1	-68.3	-5.5
Hispanic or Latino Origin											
Hispanic	76,007	97.1	22,977	97.6	1,992	25.9	52,664	93.5	0.5	-71.2	-3.6
Non Hispanic	1,175,075	97.2	178,285	97.0	27,623	29.2	216,236	91.2	-0.2	-68.0	-6.0
Race											
White Alone	1,072,144	97.2	160,932	96.9	23,471	29.5	181,779	90.9	-0.3	-67.7	-6.4
Black Alone	88,789	97.2	24,671	98.2	1,830	30.6	49,836	94.0	1.0	-66.6	-3.2
AIAN Alone	6,976	95.4	1,798	93.9	2,958	23.8	2,917	81.2	-1.5	-71.6	-14.2
Asian Alone	46,303	97.3	5,329	97.8	202	24.8	10,272	93.9	0.5	-72.5	-3.4
NHPI Alone	890	98.1	212	97.6	56	26.8	519	95.8	-0.4	-71.3	-2.3
SOR Alone	17,123	96.9	4,958	97.4	404	23.5	18,437	93.6	0.5	-73.4	-3.3
Two or more races	18,857	96.9	3,362	97.1	694	30.8	5,140	91.1	0.3	-66.0	-5.7

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

As expected, addresses in the CAPI Direct group had the lowest match ratios overall and also across race groups and Hispanic origin. This is because the CAPI Direct submode is comprised of unmailable/undeliverable and Remote Alaska addresses. On the other hand, addresses in the CATI and CAPI Nonresponse categories had match ratios that were closer in magnitude to mail match ratios. This was the case overall and across Hispanic origin and race groups, except for the AIAN CAPI Nonresponse group. In fact, addresses corresponding to the AIAN group had the lowest match ratios across all mode categories and subcategories.

Also note that address match ratios for Hispanics and non-Hispanics were close in magnitude across all modes.

APPENDIX V

Appendix V, Table 1. Person Match Ratios for Demographic Groups by Mode & CAPI Subcategory, and Differences in Match Ratios from Mail Category

2010 ACS Demographic Characteristics	Mail Match Ratios		CATI Match Ratios		CAPI Direct Match Ratios		CAPI nonresponse Match Ratios		Percentage Point Difference from Mail Match Ratio (when match ratio expressed as a % of all 2010 ACS person records)			Percentage Point Difference from Mail Match Ratio (when match ratio expressed as a % of person records w/ PIKs)		
	As a % of all person in ACS	As a % of person records w/ PIKs	As a % of all persons in ACS	As a % of person records w/ PIKs	As a % of all persons in ACS	As a % of person records w/ PIKs	As a % of all persons in ACS	As a % of person records w/ PIKs	CATI	CAPI Direct	CAPI Nonresp	CATI	CAPI Direct	CAPI Nonresp
Total	96.7	98.4	81.3	97.5	83.7	97.7	84.0	95.9	-15.4	-13.0	-12.7	-0.9	-0.7	-2.5
Hispanic or Latino Origin														
Hispanic	93.1	96.3	76.8	93.9	73.4	93.4	76.9	92.0	-16.3	-19.7	-16.2	-2.4	-3.0	-4.3
Non Hispanic	97.1	98.6	82.3	98.2	84.8	98.1	86.5	97.2	-14.8	-12.3	-10.6	-0.4	-0.5	-1.4
Race														
White Alone	97.2	98.6	81.3	97.6	84.9	97.7	84.9	96.3	-15.8	-12.2	-12.2	-1.0	-0.9	-2.3
Black Alone	96.1	98.2	84.1	98.0	81.0	98.0	86.4	97.2	-12.0	-15.2	-9.8	-0.2	-0.2	-1.0
AIAN Alone	95.9	98.2	80.8	97.4	80.3	98.9	87.3	97.6	-15.0	-15.6	-8.6	-0.8	0.7	-0.6
Asian Alone	93.7	96.8	78.0	97.2	72.9	96.1	78.5	94.7	-15.7	-20.9	-15.2	0.3	-0.8	-2.1
NHPI Alone	92.6	96.6	78.8	96.9	76.6	96.0	81.1	95.4	-13.8	-16.0	-11.6	0.3	-0.5	-1.1
SOR Alone	91.5	95.8	77.0	94.9	68.8	91.6	74.3	91.1	-14.5	-22.7	-17.2	-0.9	-4.2	-4.7
Two or more races	93.9	96.1	79.6	95.4	86.2	96.8	85.7	94.8	-14.4	-7.8	-8.2	-0.7	0.7	-1.2
Age														
0-2	79.8	82.0	67.5	79.8	68.3	80.0	70.0	79.1	-12.4	-11.6	-9.9	-2.1	-2.0	-2.8
3-17	94.3	96.2	79.6	94.7	81.8	95.5	83.1	93.8	-14.6	-12.5	-11.2	-1.5	-0.7	-2.4
18-24	96.0	98.5	81.0	97.3	82.5	97.9	82.2	96.5	-15.1	-13.6	-13.9	-1.2	-0.7	-2.0
25-44	97.4	99.2	82.1	98.3	84.9	98.9	84.3	97.4	-15.4	-12.6	-13.2	-0.9	-0.3	-1.9
45-64	98.4	99.8	83.1	99.6	85.4	99.7	88.0	99.1	-15.3	-13.0	-10.4	-0.2	-0.1	-0.6
65-74	98.5	99.8	81.8	99.8	87.0	99.9	87.0	99.5	-16.8	-11.5	-11.5	-0.1	0.0	-0.3
75+	98.6	99.9	82.8	99.8	85.7	99.8	88.4	99.5	-15.8	-12.9	-10.2	-0.1	-0.1	-0.4
Sex														
Male	96.6	98.3	81.2	97.3	83.6	97.5	83.2	95.7	-15.3	-13.0	-13.3	-0.9	-0.7	-2.6
Female	96.8	98.5	81.4	97.6	83.8	97.8	84.7	96.1	-15.4	-13.0	-12.1	-0.9	-0.6	-2.3

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

As seen in Appendix V, Table 1, person match ratios of CATI and two CAPI subcategories were close in magnitude and lower than the mail match ratio overall and across Hispanic origin and race. However, the observed differences in the match ratios across modes were mostly due to differences in PIK assignment rates between those responding by mail and those responding via CATI or CAPI.

When only person records with PIKs are taken into account, the match ratios across all modes became quite high and a lot closer in magnitude. The CAPI Nonresponse mode exhibited the lowest match ratios but the differences were not large and seemed to have been primarily driven by the Hispanic group (remember that the SOR alone group is comprised mostly of Hispanics). In fact, the person match ratios for the SOR alone and Hispanic groups remained the lowest across all mode categories and subcategories.

Finally, administrative records undercoverage of the youngest group (0-2) persisted across CAPI subcategories.

APPENDIX VI

Appendix VI, Table 1. Person-Address Match Ratios for Demographic Groups by Mode & CAPI Subcategory, and Differences from Mail Category

ACS Demographic Characteristics	Mail Match Ratios		CATI Match Ratios		CAPI Direct Match Ratios		CAPI nonresponse Match Ratios		Percentage Point Difference from Mail Match Ratio (when match ratio expressed as a % of all 2010 ACS person-address pairs)			Percentage Point Difference from Mail Match Ratio (when match ratio expressed as a % of person-address pairs w/ PIKs)		
	As a % of all person-address pairs	As a % of person-address pairs w/ PIKs	As a % of all person-address pairs	As a % of person-address pairs w/ PIKs	As a % of all person-address pairs	As a % of person-address pairs w/ PIKs	As a % of all person-address pairs	As a % of person-address pairs w/ PIKs	CATI	CAPI Direct	CAPI Nonresp	CATI	CAPI Direct	CAPI Nonresp
Total	84.8	86.2	69.9	83.8	9.8	11.5	55.0	62.8	-14.8	-74.9	-29.8	-2.4	-74.8	-23.4
Hispanic or Latino Origin														
Hispanic	77.3	80.0	64.0	78.3	5.5	7.0	50.2	60.0	-13.3	-71.8	-27.1	-1.7	-73.0	-20.0
Non Hispanic	85.4	86.8	71.1	84.9	10.3	11.9	56.7	63.7	-14.3	-75.2	-28.8	-1.9	-74.9	-23.0
Race														
White Alone	86.0	87.3	70.2	84.3	11.0	12.6	55.9	63.4	-15.8	-75.0	-30.1	-3.0	-74.7	-23.9
Black Alone	78.8	80.5	71.5	83.4	8.1	9.9	55.9	62.9	-7.3	-70.7	-23.0	2.9	-70.6	-17.6
AIAN Alone	76.4	78.3	59.9	72.1	5.1	6.2	42.3	47.3	-16.5	-71.3	-34.1	-6.2	-72.0	-31.0
Asian Alone	81.5	84.2	70.1	87.3	5.3	7.0	56.1	67.7	-11.4	-76.2	-25.4	3.1	-77.2	-16.5
NHPI Alone	73.5	76.7	64.5	79.3	6.3	7.9	53.6	63.1	-9.1	-67.2	-20.0	2.6	-68.7	-13.6
SOR Alone	74.8	78.3	64.3	79.3	4.5	6.0	48.9	60.0	-10.5	-70.3	-25.9	1.0	-72.3	-18.3
Two or more races	76.7	78.5	64.6	77.5	9.6	10.8	52.1	57.7	-12.1	-67.1	-24.6	-1.0	-67.7	-20.8
Age														
0-2	62.2	63.9	50.3	59.5	3.7	4.4	38.2	43.1	-11.9	-58.5	-24.1	-4.4	-59.5	-20.7
3-17	76.5	78.0	62.4	74.2	4.9	5.7	49.7	56.1	-14.1	-71.6	-26.8	-3.9	-72.3	-21.9
18-24	73.8	75.7	67.8	81.4	7.0%	8.3	45.9	54.0	-6.0	-66.8	-27.8	5.8	-67.4	-21.7
25-44	84.9%	86.4	72.5	86.8	9.6%	11.2	58.0	67.0	-12.4	-75.3	-26.9	0.4	-75.2	-19.4
45-64	90.4%	91.7	74.9	89.8	12.7%	14.8	64.2	72.4	-15.5	-77.7	-26.2	-1.9	-76.8	-19.3
65-74	91.5%	92.7	72.2	88.1	15.2%	17.4	61.1	69.8	-19.3	-76.3	-30.5	-4.6	-75.3	-22.9
75+	91.0%	92.2	72.3	87.2	15.3%	17.8	59.7	67.2	-18.6	-75.7	-31.3	-5.0	-74.3	-25.0
Sex														
Male	84.4%	85.9	69.6	83.4	9.8%	11.4	54.0	62.1	-14.8	-74.6	-30.4	-2.5	-74.4	-23.8
Female	85.1%	86.5	70.2	84.2	9.8%	11.5	55.9	63.5	-14.9	-75.3	-29.2	-2.4	-75.1	-23.1

Sources: Unweighted 2010 ACS and 2010 ACS Match Study Administrative Records Data

As seen in Appendix VI, Table 1, person-address match ratios for the CATI, CAPI Direct and CAPI Nonresponse categories were a lot lower than the match ratios of those responding by mail. The differences were higher for both CAPI subcategories, but especially for CAPI Direct. These differences from mail match ratios practically disappear for CATI, but were maintained for the CAPI subcategories when only those with PIKs were taken into account.

The much lower person-address match ratio for the CAPI direct category was mostly driven by the very low address match ratio of this mode subcategory. Remember CAPI Direct addresses are unmailable/undeliverable or in Remote Alaska. These type of addresses are very hard to link to or find in administrative records.

The differences in person-address match ratios for all modes across Hispanic origin were small. The AIAN group had a particularly low CAPI Nonresponse match ratio even when only those with PIKs were taken into account. As before, this could also be attributed to the very low address match ratio of this group.