

CARRA Working Paper Series
Working Paper Series #2014 - 14

**Coverage and Agreement of Administrative Records
and 2010 American Community Survey Demographic Data**

Renuka Bhaskar
U.S. Census Bureau

Adela Luque
U.S. Census Bureau

Sonya Rastogi
U.S. Census Bureau

James Noon
U.S. Census Bureau

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, D.C. 20233

Paper Issued: November 24, 2014

Disclaimer: This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data

Renuka Bhaskar, U.S. Census Bureau
Adela Luque, U.S. Census Bureau
Sonya Rastogi, U.S. Census Bureau
James Noon, U.S. Census Bureau

Abstract

The U.S. Census Bureau is researching possible uses of administrative records in decennial census and survey operations. The 2010 Census Match Study and American Community Survey (ACS) Match Study represent recent efforts by the Census Bureau to evaluate the extent to which administrative records provide data on persons and addresses in the 2010 Census and 2010 ACS. The 2010 Census Match Study also examines demographic response data collected in administrative records. Building on this analysis, we match data from the 2010 ACS to federal administrative records and third party data as well as to previous census data and examine administrative records coverage and agreement of ACS age, sex, race, and Hispanic origin responses. We find high levels of coverage and agreement for sex and age responses and variable coverage and agreement across race and Hispanic origin groups. These results are similar to findings from the 2010 Census Match Study.

1. Introduction

Administrative records include data collected by federal and state agencies in the course of providing services as well as data collected and compiled by third parties. The use of these data in censuses and surveys may reduce respondent burden and costs in an environment of declining response rates and budgets, and increasing survey and decennial census costs.

Administrative records may be able to assist with item imputation when people do not answer particular questions on a survey or census. The Census Bureau currently uses a variety of techniques to impute missing responses (Rothhaas et al. 2012, U.S. Census Bureau 2014). For some cases, responses are imputed based on additional information on the respondent or others in the household.¹ In other cases, the Census Bureau uses hot decks to impute data from a “nearest neighbor” with similar characteristics (Andridge and Little 2010, Farber et al. 2005, Fay 1999, Obenski et al. 2005, U.S. Census Bureau 2014). Some applications of this method rely on the assumption that individuals who live near each other share characteristics, which critics argue is problematic given the increasing diversity in the United States (Farber et al. 2005).

Administrative records could also assist when households do not respond to the decennial census or surveys. As part of a research program to assess strategies to reduce costs while maintaining quality for the 2020 Census, Census Bureau researchers are evaluating whether high quality administrative records can be used when households do not respond to initial contact attempts by the Census Bureau. The quality and coverage of person, address, demographic, and housing information in administrative records are critical to this evaluation.

In this paper, we focus on administrative records demographic information. Previous research on administrative records, Census 2000, and the 2010 Census indicates that the coverage and agreement of age and sex data is high in administrative records, whereas there is considerable variation across race and Hispanic origin groups (Bye and Judson 2004, Farber and Leggieri 2002a, Farber et al. 2005, Obenski et al. 2005, Rastogi and O’Hara 2012). We build on this research by evaluating demographic data in administrative records relative to the 2010 American Community Survey (ACS). We link individual records from the 2010 ACS with several different administrative records files to evaluate 1) the extent to which administrative records provide data on race, Hispanic origin, age, and sex for 2010 ACS respondents and 2) whether demographic responses in administrative records are in agreement with responses in the 2010 ACS.

We find that administrative records provide high levels of coverage and agreement for sex and age responses for individuals in the 2010 ACS. Our results indicate variation in coverage and agreement for race and Hispanic origin data. These findings concur with previous research and

¹ For example, if an individual does not report their sex but responds that they gave birth to children in the past twelve months, this indicates the person is female (U.S. Census Bureau 2014).

suggest that administrative records may be useful to assist in the determination of particular demographic items on surveys and for the decennial census.

In the following section, we provide background on previous research on administrative records demographic response data. We then discuss the data and methodology used in our analysis and present our findings on the coverage and agreement of administrative records demographic data relative to the 2010 ACS. Finally, we conclude by discussing the implications of these findings and propose future research.

2. Background

Following Census 2000, researchers evaluated administrative records demographic response data using the Statistical Administrative Records System, or StARS (Farber and Leggieri 2002b, Bye and Judson 2004).² These studies generally found lower counts for minority groups in StARS relative to Census 2000. There are two main reasons that contributed to this finding - coverage issues in administrative records as well as differences in how race and Hispanic origin were collected in administrative records relative to Census 2000. Many of the administrative data used in StARS did not yet follow the 1997 Office of Management and Budget (OMB) race and ethnic standards, which were implemented in Census 2000.

Compared to race and Hispanic origin, administrative records age and sex responses were more consistent relative to Census 2000 responses. Previous research indicates that while administrative records data do not cover younger age groups as well as they cover older age groups, age and sex responses in administrative records have, in general, high levels of coverage and agreement relative to decennial data (Farber and Leggieri 2002a, Farber et al. 2005, Obenski et al. 2005).

More recently, Rastogi and O'Hara (2012) evaluated administrative records demographic data relative to the 2010 Census. Compared to earlier work, this research utilized an expanded set of administrative records sources. Rastogi and O'Hara (2012) assessed the coverage of administrative records demographic data relative to the 2010 Census as well as the level of agreement of demographic data from individual federal and third party data sources relative to 2010 Census responses. They found that administrative records coverage of race data varied by race group. Administrative records had higher rates of race data coverage for White alone and Black alone populations and the lowest coverage for the Some Other Race (SOR) alone population. They found high levels of agreement in race response data for the White alone

² StARS 1999 was assembled from six administrative records sources: (1) Internal Revenue Service Individual Income Returns and (2) Information Returns, (3) Department of Housing and Urban Development Tenant Rental Assistance Certification System (4) Center for Medicare and Medicaid Services Medicare Enrollment Database, (5) Indian Health Service Patient Registration System, and (6) Selective Service System Registration System (Farber and Leggieri 2002a). In StARS 2000, and for subsequent years, an additional source file was added, (7) the Department of Housing and Urban Development Public and Indian Housing Information Center file.

population across federal and third party data sources. For the Black alone population, federal files had high levels of agreement relative to the 2010 Census, while third party data had moderate to low levels of agreement. Administrative records had moderate levels of agreement for the Asian alone population and low levels of agreement for other race groups such as the American Indian and Alaska Native (AIAN) alone, Native Hawaiian and Other Pacific Islander (NHPI) alone, and Two or More races populations.

Regarding Hispanic origin, Rastogi and O'Hara (2012) found that administrative records coverage of Hispanic origin data was higher for those who reported non-Hispanic in the 2010 Census compared to Hispanic. They also found moderate to high levels of agreement for the Hispanic population and high levels of agreement for the non-Hispanic population. Consistent with previous studies, they found high levels of coverage and agreement for age and sex.

Rastogi and O'Hara (2012) concluded that administrative records can assist with age and sex determination and help inform race and Hispanic origin determination. This paper replicates their analysis to evaluate whether their conclusions regarding coverage and agreement of administrative records demographic data apply to the ACS.

3. Data and Methodology

3.1 Administrative Records Data

In this analysis, we match 2010 ACS data to federal and third party administrative source files obtained by the Census Bureau.

The federal administrative records used in this analysis include three different files from the Department of Housing and Urban Development (HUD) for those who receive rental assistance, housing assistance and Federal Housing Authority loan application records, the Indian Health Service (IHS) Patient Registration System, Medicare Enrollment Database (MEDB), Selective Service Registration System (SSS), Temporary Assistance for Needy Families (TANF), and the Social Security Administration (SSA) Numerical Identification System (Numident) and Supplemental Security Income files. Along with these federal data, data from five third party data vendors were also used. The third party data vendors gather data from sources such as telephone book white pages, utility records, property records, and credit bureau header information (see Rastogi and O'Hara (2012) for more information on federal and third party data sources). Since agreements with third party data vendors prohibit direct comparison of data across sources, third party file names are not used. Instead, third party data files are called third party file 1, third party file 2, etc. throughout this paper.

Finally, in addition to federal and third party administrative records, we use previous census records which include data from Census 2000 and ACS 2001 to 2009. These datasets were

combined and the most recent response for an individual for each demographic item (e.g., race or age) was utilized.³

Administrative records sources vary in terms of the demographic data they provide. Some of the files, such as the SSS file, do not contain information on all of the demographic variables included in our analysis. This file includes information on age and sex, but not on Hispanic origin and race. Other files, such as those from HUD, contain data on all four of the demographic variables in this analysis.

Additionally, these data do not uniformly collect and report data on Hispanic origin and race. First, there are differences in the categories that are available. Unlike the Census Bureau, the HUD and TANF files do not include a category for SOR. The IHS file differs in that it only identifies persons as either AIAN or non-AIAN. The Numident and MEDB, based on the OMB's 1977 race and ethnicity standards, have a combined category for Asian and Pacific Islander.⁴ For these two files, we evaluate agreement of race data by comparing the combined Asian and Pacific Islander category to the ACS Asian alone and NHPI alone categories separately.⁵

Second, there are differences in how federal agencies handle race and Hispanic origin and multiple race reporting. Except for the Numident and MEDB, all of the federal data we use follow OMB's 1997 race and ethnicity standards which define race and Hispanic origin as two separate concepts and prescribes that agencies should allow respondents to choose more than one race when separate Hispanic origin and race questions are used.⁶ The Numident and MEDB treat Hispanic origin and race as one concept and only have single race responses. Additionally, prior to 1980, "Hispanic" was not a category on these files, individuals were coded as "White", "Black", or "Other".

When we use the Numident and MEDB files, individuals whose race was listed as Hispanic were coded as Hispanics with missing race. Individuals whose race was listed as a race other than

³ Note that we only include responses that did not require editing or imputation.

⁴ In the Numident, the SSA collected race data from 1936 to 1980 via the Social Security application based on three categories: "White," "Black," and "Other." In 1980, SSA changed its categories to "White," "Black," "Hispanic," "Asian, Asian American, or Pacific Islander," and "American Indian or Alaskan Native." The SSA then stopped collecting race and ethnic data for the vast majority of people who obtain Social Security Numbers when it transitioned to the Enumeration at Birth system in 1989 (Scott 1999). A large proportion of race and ethnic data in MEDB comes from the SSA. MEDB used other sources to update race and ethnic information, and imputation methods are applied when race and ethnic data are missing (McBean 2006).

⁵ Rastogi and O'Hara (2012) calculated Numident-2010 Census agreement for Asians and NHPI using this same methodology. For the MEDB file, the codebook indicates that there is only an Asian category and does not indicate that Pacific Islanders are included in this category. Based on this information, Rastogi and O'Hara (2012) calculated the agreement between the Asian category in MEDB and those who reported Asian in the 2010 Census. However, studies indicate that the "Asian" category in the MEDB file combines Asian and Pacific Islanders, thus our approach is to treat MEDB in the same way we evaluate Asians and Pacific Islanders in the Numident (Eicheldinger and Bonito 2008, Zaslavsky et al. 2012).

⁶ When collecting and tabulating data on race and ethnicity, federal agencies must adhere to guidance from the OMB's 1997 *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. The standards are available online at <http://www.whitehouse.gov/omb/fedreg/1997standards.html>.

Hispanic were coded as that race group with missing Hispanic origin information. Thus, we only coded individuals as Hispanic in the Numident and MEDB files. The rest of the individuals were coded as missing Hispanic origin. This methodology differs from research by Rastogi and O'Hara (2012) in which a race response other than Hispanic was coded as non-Hispanic. Because the SSA did not include the category "Hispanic" until 1980, the methodology used by Rastogi and O'Hara (2012) resulted in the assignment of "non-Hispanic" to all individuals whose data was collected prior to 1980, as "Hispanic" was not a category until this time.

The third party files have detailed ethnic origin data that is modeled using surnames and geography that we coded into race and Hispanic origin categories. These ethnic origin variables treat race and Hispanic origin as one concept and also do not allow for multiple responses. Thus, individuals coded as Hispanic do not have a race. Unlike with the Numident and MEDB, individuals in third party data who are not coded as Hispanic but who have other race or origin information are coded as non-Hispanic.⁷ We coded the third party files differently from the Numident and MEDB files because "Hispanic" was not a category until 1980 in the Numident and MEDB files, thus many Hispanic individuals were coded as "Other" in those files. In third party data, "Hispanic" is a category, thus we coded individuals with a non-Hispanic race or origin as non-Hispanic (rather than coding those with an "Other" race or origin as non-Hispanic).

Most of the files included in this analysis contain information on date of birth. However, some files do not contain full day, month, and year information. One file from HUD provides only year of birth, and two of the third party files contain only year and month.

3.2 2010 ACS

The ACS is an ongoing survey that samples a small percentage of the U.S. population each year on a variety of socioeconomic, demographic, and housing topics. For the purposes of this analysis, we focus on 2010 ACS age, sex, race, and Hispanic origin responses. Data for the 2010 ACS were collected from January through December of 2010.

For this analysis, we include both persons in housing units and group quarters in the fifty states and the District of Columbia.⁸ Data on persons living in housing units were collected by three modes – mail, computer assisted telephone interview (CATI), or computer assisted personal interview (CAPI). For the population living in group quarters which include prisons, dormitories, and nursing homes, data are collected through a group quarters visit.

⁷ There was a small error in Rastogi and O'Hara (2012) regarding Hispanic origin coding for third party data. In their study, individuals with no race or origin information in third party files were coded as non-Hispanic, and in our analysis these individuals are coded as missing Hispanic origin information. This difference does not have a large impact, and our results for third party coverage and agreement of Hispanic origin data are similar to what they found.

⁸ Puerto Rico was not included in our analysis.

This paper uses unweighted ACS response data for the 4.5 million individuals in the 2010 person response file. Weights are not used because this study is assessing administrative demographic response data for persons in the 2010 ACS, not using the ACS to evaluate the demographic information provided by administrative records for the total U.S. population.

During the editing process for the ACS, when a response to a particular question is incomplete or invalid, the Census Bureau may have to impute a response. In some cases, the imputed response can be assigned based on other information for the respondent or another respondent in the household. In other cases, responses are allocated using a hot deck matrix in which statistical methods are used to supply responses based on data from other people in the sample (U.S. Census Bureau 2014). Allocation rates for the demographic items studied here (age, sex, race, and Hispanic origin) ranged from 0.1 to 1.8 in the 2010 ACS.⁹ Some parts of our analysis focus on all ACS respondents, whereas other sections differentiate on whether they required imputation or allocation.

3.3 Methodology

The Census Bureau's Person Identification Validation System (PVS) facilitates the unduplication of records within administrative records files and allows for record linkage across different data sources (such as surveys, censuses and administrative records). The PVS assigns unique person identifiers called Protected Identification Keys (PIKs) to person records in administrative records, survey or census files by comparing personally identifiable information (PII) such as name, date of birth, and address contained in those files to PII on person reference files.¹⁰ Once PIKs are assigned, different data sources can be linked or matched using these identifiers. For this study, we linked administrative records data to the 2010 ACS. After PIKs are assigned, PII is removed from the data to anonymize the data and preserve confidentiality so it can be used for statistical purposes and research. For more information on the PVS, see Wagner and Layne (2014).

We first evaluate administrative records coverage of demographic data for the 2010 ACS sample. Administrative records data sources are compiled into a composite file and variables are created to indicate the presence of demographic data. Persons from the 2010 ACS are then matched to the administrative records composite by PIK. For a given demographic group, coverage is defined as the number of PIKs in the ACS that have administrative records demographic data

⁹ ACS item allocation rates can be found at https://www.census.gov/acs/www/methodology/item_allocation_rates_data/

¹⁰ Reference files contain information from the SSA enhanced with address data obtained from federal administrative record files. Reference files contain all variants of a person's name, date of birth, and sex, as well as current and recent addresses.

divided by the number of persons in the ACS.¹¹ We report coverage by demographic group, survey mode, and administrative records data source.

After discussing the coverage of demographic data, we present findings on the agreement of demographic responses from administrative records and the 2010 ACS. Specifically, we evaluate the response agreement between the 2010 ACS and administrative records for 2010 ACS respondents that match to administrative records and for whom demographic response information is available in the administrative records. For this portion on the analysis, we focus on 2010 ACS respondents whose response to each demographic item was not imputed.

4. Results

4.1 Coverage of Demographic Response Data

Table 1 shows whether demographic data for 2010 ACS persons are present in any of the administrative records files for each demographic variable. Note that this does not indicate that the demographic information is in agreement, just that there is a response available. Of the almost 4.5 million persons in the 2010 ACS, administrative records provide Hispanic origin responses for approximately 3.6 million, or 79.9 percent. Coverage of Hispanic origin data is higher for those who reported non-Hispanic in the 2010 ACS (81.9 percent) relative to Hispanic (65.9 percent).

Data on race are available for approximately 3.6 million or 80.8 percent of 2010 ACS person records, though this varies by race group. Administrative records provide the highest level of race data coverage for those who reported Black alone (82.8 percent) and White alone (82.7 percent) in the 2010 ACS. Coverage is lowest (45.7 percent) for those who reported SOR alone.

About 93 percent of people (4.2 million) in the 2010 ACS had age data in administrative records. The results are similar for sex data. Coverage is high for all age groups, ranging from 91.0 percent to 95.4 percent. Coverage for females, 93.2 percent, is slightly higher than for males, 92.7 percent.

Our findings for race, age, and sex are consistent with those found by Rastogi and O'Hara (2012). However, as mentioned earlier, our analysis differs from theirs in terms of the coding of Hispanic origin data from certain administrative records files, and as expected, we find lower coverage of Hispanic origin data for both Hispanics and non-Hispanics. For the other variables, there is higher demographic coverage for persons in the 2010 ACS relative to the 2010 Census. This is expected as Luque and Bhaskar (2014) found that administrative records provide higher coverage of person records for the 2010 ACS relative to the 2010 Census.

¹¹ Because PIKs are required to be in our matching universe, only individuals with PIKs can be included in our numerator, while all individuals in the 2010 ACS are included in the denominator for calculating coverage. Approximately 94.2 percent of individuals in the 2010 ACS were assigned a PIK (Luque and Bhaskar 2014).

4.1.1. Coverage by Mode of Data Collection

The next series of tables focus on coverage of each demographic item by ACS mode of data collection. For each of the demographic variables included in this analysis, coverage is higher for those collected through mail compared to those collected through CATI or CAPI. This is likely due in part to lower PIK and person coverage rates for CATI and CAPI respondents, as found by previous research (Luque and Bhaskar 2014).¹²

Table 2 shows whether Hispanic origin data are present in administrative records for each mode of ACS data collection. Across modes, non-Hispanic coverage ranges from 71.1 percent in CAPI to 86.1 percent in Mail. For Hispanics, the range is 56.0 percent in CAPI to 76.1 percent in Mail. For all modes, coverage of Hispanic origin data is higher for non-Hispanics than for Hispanics.

Table 3 shows whether race data are present in administrative records for each mode. For all groups, coverage rates are lowest in CATI or CAPI. For most race groups, coverage rates are highest for those responding via mail. One exception is the Two or More Races group, for which the highest coverage rates are in group quarters visit. For all modes of data collection, coverage is lowest for those whose 2010 ACS race is SOR alone. Those who reported Black alone had either the highest or second highest coverage across all modes.

Table 4 shows whether age data are present in administrative records for each mode and age group. Once again, administrative records provide the highest coverage of age data for 2010 ACS respondents who responded via mail. Among those who responded via mail, there is little variation by age group. For those in group quarters, we see much more variation in age data coverage, which may be due in part to small sample sizes for the 0 to 2 and 3 to 17 age groups.

Table 5 shows whether sex data are present in administrative records for each mode. Similar to other demographic characteristics, administrative records coverage of 2010 ACS respondents is highest for those responding by mail and lowest for those in CATI and CAPI. There is generally little variation between sexes within all response modes.

Our findings of coverage of demographic data by survey mode are not unexpected given the findings from previous research using administrative records and the 2010 Census. For all demographic variables, coverage was highest for responses collected through mail mode (Rastogi and O'Hara 2012). Coverage for persons in the 2010 Census who responded through nonresponse followup operations are comparable to our results for 2010 ACS respondents in CATI and CAPI. Patterns by demographic group within modes are also comparable. Once again, with the exception of Hispanic origin, we find coverage rates for persons in the 2010 ACS to be in general slightly higher than those found in analysis using the 2010 Census.

¹² Individuals who responded to the 2010 ACS by mail had a PIK assignment rate of 98.3 percent while those who responded via CATI and CAPI had lower PIK rates of 83.4 percent and 87.4 percent respectively. Administrative records coverage of persons in the 2010 ACS who received PIKs was 98.4 percent for those responding by mail, 97.5 percent for CATI, and 96.1 percent for CAPI (Luque and Bhaskar 2014).

4.1.2 Coverage by Administrative Records Data Source

The next series of tables evaluate administrative records coverage of 2010 ACS demographic response data for each of the administrative records sources used in this analysis. We find great variation in coverage by the different administrative records sources, which is expected given the variation in the files in terms of size and populations they cover. For all demographic items, the combined federal data provide higher coverage relative to combined third party data.

Table 6 presents administrative records coverage of 2010 ACS Hispanic origin data by federal and third party data sources. Previous census records provide the highest coverage for both Hispanics (47.3 percent) and non-Hispanics (73.2 percent), and the Numident provides the second highest coverage for Hispanics. Note that the Numident and the MEDB file can only indicate that an individual is Hispanic, and they cannot be used to determine that a person is non-Hispanic. That is, Hispanic origin data take the value of either ‘Hispanic’ or missing in the Numident and MEDB files. Individuals identified as non-Hispanic in the 2010 ACS are, then, likely to have missing Hispanic origin data in the Numident and MEDB. This results in a much higher coverage for Hispanics relative to non-Hispanics by these files. The remaining federal files provide minimal coverage of Hispanic origin data. Coverage by third party files ranges from 13.5 to 36.5 percent for Hispanics and 21.7 to 59.8 percent for non-Hispanics.

Table 7 shows administrative records coverage of race response data by federal and third party data sources. For most individual source files, coverage is highest for the White alone or Black alone groups and is generally low for the SOR alone group. One exception is the IHS file, which as expected has low coverage of most race groups but notable coverage (52.1 percent) of those who report AIAN alone. For all race groups however, the source files with the highest coverage are previous census records or the Numident file.

Table 8 shows administrative records coverage of 2010 ACS age response data by federal and third party data sources. As discussed earlier, administrative records provide high coverage of age data overall. Here we see that federal data provide higher coverage relative to third party data, particularly for younger age groups. Table 8 indicates that much of the age data comes from the Numident file which has high coverage (over 90 percent) for respondents of all ages in the 2010 ACS. As expected, the MEDB file, containing Medicare enrollee data, also provides high coverage for the oldest two age groups.

Table 9 shows administrative records coverage of 2010 ACS sex response data by federal and third party data sources. The Numident file provides the highest coverage of sex data for both males (92.5 percent) and females (93.1 percent), followed by data from previous census records (70.1 percent for males and 72.4 percent for females).

These results by source file are generally similar to what was found by Rastogi and O’Hara (2012), indicating that the files that provide high coverage for each demographic item for persons in the 2010 Census also provide high coverage in the 2010 ACS. For Hispanic origin,

coverage rates for the Numident and MEDB are notably lower than found by Rastogi and O'Hara due to the differences in methodology described in the data section.

4.1.3 Coverage of Allocated ACS responses

Thus far, we have presented coverage of demographic response data for all respondents in the 2010 ACS whether their response was allocated or not. As mentioned earlier, for cases when a response to a demographic item is incomplete or invalid, the Census Bureau allocates a response based on additional information. In Table 10, we present allocation rates for each of the variables we include in our analysis – Hispanic origin, race, age, and sex. For all variables, allocation rates are low. Hispanic origin had the highest number of responses allocated (96,651), however this represents only 2.2 percent of the 4.5 million respondents to the 2010 ACS.¹³

Table 10 also shows the coverage of demographic response data for persons in the 2010 ACS whose ACS response was allocated. These coverage rates, which range from 48.7 percent for age to 69.4 percent for sex, are lower than the overall coverage rates for the 2010 ACS sample shown in Table 1.

4.2 Agreement of Demographic Response Data

For the next section of the paper, we focus on those individuals in the 2010 ACS who have demographic response data in administrative records and evaluate whether their administrative records demographic responses agree with their 2010 ACS demographic responses. Only cases for which the 2010 ACS demographic item did not require editing or imputation are considered.

Table 11 shows the agreement in Hispanic origin data between the 2010 ACS and each of the administrative records sources included in this analysis. As shown in the first row, there are 255,227 Hispanics in the 2010 ACS for whom previous census records provide Hispanic origin information. Of these individuals, 233,699 (91.6 percent) had the same Hispanic origin response in both the 2010 ACS and in previous census records.

Agreement rates for the Numident and MEDB file are not shown in the table but are discussed in the notes below the table. As described earlier, the Numident and MEDB include a Hispanic category, but non-Hispanic origin cannot be determined in these files. When we compare individuals who are Hispanic in the 2010 ACS to those who have Hispanic origin information in the Numident or MEDB, we are by default comparing only those who reported Hispanic in both data sources, which would result in 100 percent agreement. If we included those with missing Hispanic origin information in the Numident and MEDB, agreement rates for 2010 ACS

¹³ The allocation rates shown in Table 10 apply to the unweighted 2010 ACS. Allocation rates for the weighted data range can be found at: https://www.census.gov/acs/www/methodology/item_allocation_rates_data/

Hispanics would be 52.9 percent for the Numident and 26.9 percent for MEDB. We do not show agreement for non-Hispanics for these files because while there are individuals who reported non-Hispanic in the 2010 ACS and are present in the Numident and MEDB, these files do not include a non-Hispanic category. With the exception of the Numident and MEDB, agreement of Hispanic origin data is higher for non-Hispanics relative to Hispanics. Agreement for non-Hispanic response is high (98.0 percent or higher) and consistent across source files. However, there is much more variation for individuals that reported Hispanic origin in the 2010 ACS. For these, agreement in responses range from 68.9 to 91.6 percent.

Table 12 shows the agreement in race data between the 2010 ACS and each of the federal and third party administrative data sources. This percentage varies a great deal by race and source file. In general, the agreement of race data is highest for those whose 2010 ACS race is White alone, with over 94 percent of administrative records responses from each data source in agreement with the ACS race response.¹⁴ When looking at federal files, agreement for those whose ACS race is Black alone is also high (over 88 percent); however, agreement of race responses from third party data is much lower. In fact, this is the case for most individuals except those whose 2010 ACS race was White alone or Asian alone. The lower agreement rates from third party data are expected given that third party files model race data.

As already mentioned in the data section, many administrative files do not include a category for SOR or allow for multiple race options. For files that do include these response options, agreement is consistently low. Specifically, the percent of race responses in agreement range from 1.3 to 53.6 percent for the SOR alone group, and from 2.5 to 35.0 percent for the Two or More Races group. Note that for the Two or More Race category, we are evaluating whether persons have multiple races reported in both the administrative records file and the 2010 ACS, but we are not evaluating whether those multiple races are the same. As described in the data section, the Numident and MEDB files have a combined Asian and Pacific Islander category. In Table 12, this combined group is compared separately to Asians and to NHPI in the 2010 ACS. For the Numident, the percent of responses in agreement is 82.9 percent for those whose 2010 ACS race is Asian alone and 72.9 percent for those whose 2010 ACS race is NHPI alone. For MEDB, agreement percentages are 55.7 percent for Asian alone and 42.0 percent for NHPI alone.

For most race groups, the Numident or previous census records provide race response data with the highest percentage agreement. The IHS data provide responses that are in high agreement with the 2010 ACS for those whose 2010 ACS race is AIAN alone (97.9 percent).

Next, we evaluate the agreement of administrative records date of birth data. Table 13 shows the percent of 2010 ACS responses for which administrative records day, month, and year of birth are in agreement. Agreement of date of birth relative to responses in the 2010 ACS is generally

¹⁴ For two files from the Department of Housing and Urban Development, there is no statistical difference in agreement rates for 2010 ACS respondents who were White alone or Black alone.

high among federal files (over 90 percent for many files and age groups) and lower among third party files (under 80 percent for most files and age groups).

This table only includes administrative records files that contain information on day, month, and year of birth. Some files do not provide data on all three components of the birth date but do provide some information. For example, the HUDCHUMS¹⁵ file only includes information on year of birth, not day or month, and the year of birth is in high agreement (over 96 percent for all age groups) with ACS year of birth responses. Similarly, third party files 5 and 7 only include data on month and year of birth. Agreement of month and year of birth between these files and ACS responses range from 73.5 to 85.5 and 70.2 to 82.1 respectively by age group for those over the age of 18.

Finally, we look at the agreement of sex data in administrative records and the 2010 ACS. As shown in Table 14, administrative records provide high quality sex data, with agreement rates of over 94 percent for both sexes across all federal and third party data sources. Focusing on federal files, agreement for 2010 ACS sex responses ranges from 97.9 to 100.0 percent.

Comparisons of these results to those found by Rastogi and O'Hara (2012) must be considered with caution, but it is useful to note that the results are generally similar. For Hispanic origin and race, agreement patterns for most source files are similar to Rastogi and O'Hara's (2012) results. One exception is Hispanic origin agreement for the Numident and MEDB files where our analysis used a different methodology. In our study, we evaluate agreement of date of birth data, which is slightly different from Rastogi and O'Hara's (2012) evaluation of agreement of age. However, both analyses indicate that federal administrative records provide age and date of birth information that is in high agreement with the 2010 Census and 2010 ACS. For working age adults, our analysis finds lower date of birth agreement rates for third party files compared to Rastogi and O'Hara's (2012) analysis of age agreement. Finally, for agreement of sex responses, our findings are very close, differing in most cases by less than one percentage point.

5. Conclusions and Future Research

This analysis finds that administrative records contain demographic response data for the majority of persons in the 2010 ACS with variation by demographic group, data source, and ACS mode of data collection. Overall, coverage of age and sex data is high across demographic groups. Coverage of Hispanic origin and race data is more variable, but for some groups, administrative records provide high coverage. Coverage is higher for persons who respond to the 2010 ACS by mail and for certain administrative files. For all demographic items, coverage is lower for persons whose 2010 ACS response required allocation.

¹⁵ Computerized Homes Underwriting Management Systems file from the Department of Housing and Urban Development.

Furthermore, some administrative records sources, especially the Numident and previous census records, provide demographic response data that are in agreement with 2010 ACS responses, particularly for sex. Date of birth responses from federal files are also in high agreement relative to the 2010 ACS. The results for race and Hispanic origin are more variable, though for some groups, certain files do provide data that are in high agreement with 2010 ACS responses. These results concur with Rastogi and O'Hara's (2012) findings, suggesting that administrative records may be useful to assist in the determination of particular demographic items.

Further research will help determine how the Census Bureau can utilize demographic information from administrative records for operational purposes. For example, research is underway to assess lower agreement rates for particular race and Hispanic origin groups. Researchers at the Census Bureau are currently evaluating how to determine a single demographic response when demographic data across administrative records are discrepant. Furthermore, we are evaluating race and Hispanic origin responses from additional administrative data sources such as the Supplemental Nutrition Assistance Program and the Medicaid Statistical Information System.

Additionally, our analysis of agreement for all demographic responses focused on those with non-allocated demographic responses in the 2010 ACS. Future work should compare demographic responses from administrative records to responses derived from allocation procedures for individuals in the ACS to determine how demographic responses in administrative records compare to those derived through current imputation procedures.

Finally, as we move forward toward the 2020 Census, this analysis should be replicated using future years of the ACS to determine if and how patterns may change over time as administrative records systems and the ACS evolve throughout the decade.

References

- Andridge, Rebecca R. and Roderick J.A. Little. 2010. "A Review of Hot Deck imputation for Survey Non-response," *International Statistical Review* 78(1):40-64.
- Bye, Barry and Dean Judson. 2004. "Results From the Administrative Records Experiment in 2000," *Census 2000 Synthesis Report No. 16*, U.S. Census Bureau.
- Eicheldinger, Celia and Arthur Bonito. 2008. "More Accurate Racial and Ethnic Codes for Medicare," *Health Care Financing Review* Volume 29, Number 3.
- Farber, James and Charlene Leggieri. 2002a. "Building and Validating a National Administrative Records Database for the United States," *New Zealand Conference on Database Integration*.
- Farber, James and Charlene Leggieri. 2002b. "Building and Validating the Prototype Statistics Administrative Records System," *Administrative Records Memorandum Series*, U.S. Census Bureau.
- Farber, James, Deborah Wagner, and Dean Resnick. 2005. "Using Administrative Records for Imputation in the Decennial Census," *Proceedings of the 2005 Joint Statistical Meetings*, Survey Research Methods Section, Alexandria, VA: American Statistical Association.
- Fay, Robert E. 1999. "Theory and Application of Nearest Neighbor Imputation in Census 2000," *Proceedings for the Section on Survey Research Methods American Statistical Association*, 112-121.
- Luque, Adela and Renuka Bhaskar. 2014. "2010 American Community Survey Match Study," *Center for Administrative Records Research and Applications Working Paper #2014-03*, U.S. Census Bureau.
- McBean, A. Marshall. 2006. "Improving Medicare's Data on Race and Ethnicity," *National Academy of Social Insurance Medicare Brief No 15*.
- Obenski, Sally, James Farber, and Gary Chappell. 2005. "Research to Improve Census Imputation Methods: Item Results and Conclusions," *Proceedings of the 2005 Joint Statistical Meetings*, Survey Research Methods Section, Alexandria, VA: American Statistical Association
- 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.
- Rothhaas, Cynthia, Frederic Lestina, and Joan M. Hill. 2012. "2010 Decennial Census: Item Nonresponse and Imputation Assessment Report," *2010 Census Program for Evaluations and Experiments, 2010 Planning Memoranda Series No. 173*, U.S. Census Bureau.

Scott, Charles G. 1999. "Identifying the Race or Ethnicity of SSI Recipients," Social Security Bulletin. Vol. 62 No 4.

U.S. Census Bureau. 2014. "Design and Methodology: American Community Survey," U.S. Government Printing Office, Washington, DC.

Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," Center for Administrative Records Research and Applications Working Paper #2014-01, U.S. Census Bureau.

Zaslavsky, Alan M., John Z. Ayanian, and Lawrence B. Zaborski. 2012. "The Validity of Race and Ethnicity Enrollment Data for Medicare Beneficiaries," Health Service Research 47 (3 Pt 2): 1300-1321.

Acknowledgements

We would like to thank Amy O'Hara, Ciara Nugent, and Karen King for their valuable comments and suggestions.

Table 1. Administrative Records Coverage of 2010 ACS Demographic Data

ACS Demographic Characteristics	2010 ACS	Coverage by Administrative Records Demographic Response Data	
		Number	Percent
Hispanic or Latino Origin	4,470,984	3,570,994	79.9
Hispanic	559,937	369,107	65.9
Not Hispanic	3,911,047	3,201,887	81.9
Race	4,470,984	3,612,282	80.8
White Alone	3,515,419	2,908,898	82.7
Black Alone	445,616	369,107	82.8
American Indian or Alaska Native Alone	43,607	35,721	81.9
Asian Alone	193,382	147,320	76.2
Native Hawaiian and Other Pacific Islander Alone	6,192	4,673	75.5
Some Other Race Alone	152,031	69,437	45.7
Two or More Races	114,737	77,126	67.2
Age	4,470,984	4,171,228	93.3
0 to 2	152,793	142,536	93.3
3 to 17	852,284	789,969	92.7
18 to 24	385,288	350,487	91.0
25 to 44	1,079,967	987,784	91.5
45 to 64	1,279,630	1,212,902	94.8
65 to 74	383,305	365,524	95.4
75 and older	337,717	322,026	95.4
Sex	4,470,984	4,156,195	93.0
Male	2,180,441	2,020,635	92.7
Female	2,290,543	2,135,560	93.2

Sources: Unweighted 2010 ACS and Administrative Records Data

Note: As mentioned in the text, individuals who were not assigned a Protected Identification Key are not included in our universe for matching to administrative records. These individuals are, however, included in the denominator. This applies to all coverage tables in this paper.

Table 2. Administrative Records Coverage of 2010 ACS Hispanic Origin Data by Mode of Data Collection

2010 ACS Survey Mode and Hispanic Origin	2010 ACS	Administrative Records	
		Number	Percent
Mail	2,972,654	2,534,708	85.3
Hispanic	250,798	190,910	76.1
Not Hispanic	2,721,856	2,343,798	86.1
CATI	545,695	379,617	69.6
Hispanic	93,338	54,549	58.4
Not Hispanic	452,357	325,068	71.9
CAPI	807,687	544,223	67.4
Hispanic	198,754	111,372	56.0
Not Hispanic	608,933	432,851	71.1
Group Quarters Visit	144,948	112,446	77.6
Hispanic	17,047	12,276	72.0
Not Hispanic	127,901	100,170	78.3

Sources: Unweighted 2010 ACS and Administrative Records Data

Table 3. Administrative Records Coverage of 2010 ACS Race Data by Mode of Data Collection

ACS Survey Mode and Race	2010 ACS	Administrative Records	
		Number	Percent
Mail	2,972,654	2,568,517	86.4
White Alone	2,466,367	2,165,031	87.8
Black Alone	213,500	188,467	88.3
American Indian or Alaska Native Alone	18,954	16,462	86.9
Asian Alone	139,706	112,276	80.4
Native Hawaiian and Other Pacific Islander Alone	2,994	2,447	81.7
Some Other Race Alone	58,411	33,333	57.1
Two or More Races	72,722	50,501	69.4
CATI	545,695	376,995	69.1
White Alone	422,520	293,695	69.5
Black Alone	64,734	49,689	76.8
American Indian or Alaska Native Alone	4,997	3,512	70.3
Asian Alone	18,692	12,404	66.4
Native Hawaiian and Other Pacific Islander Alone	867	597	68.9
Some Other Race Alone	19,259	8,564	44.5
Two or More Races	14,626	8,534	58.3
CAPI	807,687	543,977	67.3
White Alone	529,963	366,350	69.1
Black Alone	134,177	102,290	76.2
American Indian or Alaska Native Alone	17,935	14,306	79.8
Asian Alone	30,843	19,891	64.5
Native Hawaiian and Other Pacific Islander Alone	2,058	1,407	68.4
Some Other Race Alone	69,311	24,986	36.0
Two or More Races	23,400	14,747	63.0
Group Quarters Visit	144,948	122,793	84.7
White Alone	96,569	83,822	86.8
Black Alone	33,205	28,661	86.3
American Indian or Alaska Native Alone	1,721	1,441	83.7
Asian Alone	4,141	2,749	66.4
Native Hawaiian and Other Pacific Islander Alone	273	222	81.3
Some Other Race Alone	5,050	2,554	50.6
Two or More Races	3,989	3,344	83.8

Sources: Unweighted 2010 ACS and Administrative Records Data

Table 4. Administrative Records Coverage of 2010 ACS Age Data by Mode of Data Collection

ACS Survey Mode and Age	2010 ACS	Administrative Records	
		Number	Percent
Mail	2,972,654	2,909,741	97.9
0 to 2	94,376	91,773	97.2
3 to 17	529,132	515,265	97.4
18 to 24	208,242	201,594	96.8
25 to 44	685,353	666,781	97.3
45 to 64	918,814	904,802	98.5
65 to 74	293,368	289,397	98.6
75 and older	243,369	240,129	98.7
CATI	545,695	447,126	81.9
0 to 2	16,851	14,212	84.3
3 to 17	121,609	100,687	82.8
18 to 24	37,210	30,049	80.8
25 to 44	114,651	91,024	79.4
45 to 64	163,597	135,592	82.9
65 to 74	48,516	39,704	81.8
75 and older	43,261	35,858	82.9
CAPI	807,687	685,809	84.9
0 to 2	41,438	36,449	88.0
3 to 17	197,667	171,334	86.7
18 to 24	90,548	74,364	82.1
25 to 44	245,013	199,964	81.6
45 to 64	175,169	153,020	87.4
65 to 74	34,647	30,210	87.2
75 and older	23,205	20,468	88.2
Group Quarters Visit	144,948	128,552	88.7
0 to 2	128	102	79.7
3 to 17	3,876	2,683	69.2
18 to 24	49,288	44,480	90.2
25 to 44	34,950	30,015	85.9
45 to 64	22,050	19,488	88.4
65 to 74	6,774	6,213	91.7
75 and older	27,882	25,571	91.7

Sources: Unweighted 2010 ACS and Administrative Records Data

Table 5. Administrative Records Coverage of 2010 ACS Sex Data by Mode of Data Collection

ACS Survey Mode and Sex	2010 ACS	Administrative Records	
		Number	Percent
Mail	2,972,654	2,898,947	97.5
Male	1,429,995	1,393,748	97.5
Female	1,542,659	1,505,199	97.6
CATI	545,695	446,372	81.8
Male	261,572	213,765	81.7
Female	284,123	232,607	81.9
CAPI	807,687	682,517	84.5
Male	399,824	335,213	83.8
Female	407,863	347,304	85.2
Group Quarters Visit	144,948	128,359	88.6
Male	89,050	77,909	87.5
Female	55,898	50,450	90.3

Sources: Unweighted 2010 ACS and Administrative Records Data

Table 6. Administrative Records Coverage of 2010 ACS Hispanic Origin Data by Administrative Records Source File

Administrative Records Source File	Hispanic		Not Hispanic	
	Number	Percent	Number	Percent
ACS 2010	559,937	100.0	3,911,047	100.0
Federal Files	349,157	62.4	2,924,191	74.8
Previous Census Records	264,951	47.3	2,862,688	73.2
Numident	237,519	42.4	11,245	0.3
HUDCHUMS	21,434	3.8	142,259	3.6
HUDPIC	15,251	2.7	68,140	1.7
HUDTRACS	270	0.0	1,810	0.0
MEDB	11,399	2.0	537	0.0
TANF	3,692	0.7	24,911	0.6
Third Party Files	238,462	42.6	2,611,540	66.8
Third Party File 1	138,764	24.8	1,715,603	43.9
Third Party File 2	204,264	36.5	2,339,060	59.8
Third Party File 3	156,251	27.9	1,982,510	50.7
Third Party File 4	75,402	13.5	847,371	21.7

Sources: Unweighted 2010 ACS and Administrative Records Data

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

TANF: Temporary Assistance for Needy Families file

Table 7. Administrative Records Coverage of ACS 2010 Race Data by Administrative Records Source File

Administrative Records Source File	White Alone	Black Alone	American Indian or Alaska Native Alone	Asian Alone	Native Hawaiian or Other Pacific Islander Alone	Some Other Race Alone	Two or More Races
2010 ACS	3,515,419	445,616	43,607	193,382	6,192	152,031	114,737
Federal Files	82.2	82.2	81.4	74.2	74.6	43.9	65.9
Previous Census Records	74.5	62.1	60.5	52.7	48.7	37.5	53.3
Numident	68.9	71.9	60.9	66.5	63.8	9.6	42.5
HUDCHUMS	3.5	3.9	2.0	1.8	3.5	2.2	2.2
HUDPIC	1.0	8.2	2.2	1.3	3.7	3.0	3.5
HUDTRACS	0.4	2.1	0.6	0.6	0.4	0.7	0.9
IHS	0.1	0.0	52.1	0.0	-	0.1	4.1
MEDB	19.7	15.2	11.7	9.8	8.1	3.8	7.2
TANF	0.4	1.9	2.2	0.2	3.2	0.5	1.6
Third Party Files	62.8	55.5	39.8	49.4	32.1	8.1	33.9
Third Party File 1	42.0	32.8	17.7	29.6	14.3	3.3	17.8
Third Party File 2	54.5	49.3	32.2	40.2	23.6	6.2	28.1
Third Party File 3	47.4	36.6	26.3	32.2	19.7	3.9	22.1
Third Party File 4	19.9	18.3	11.8	13.7	9.2	1.9	10.7

Sources: Unweighted 2010 ACS and Administrative Records Data

Note: A dash "-" indicates that the cell is suppressed for disclosure avoidance purposes

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

IHS: Indian Health Service

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

TANF: Temporary Assistance for Needy Families file

Table 8. Administrative Records Coverage of 2010 ACS Age Data by Administrative Records Source File

Administrative Records Source File	Age						
	0 to 2	3 to 17	18 to 24	25 to 44	45 to 64	65 to 74	75 and older
ACS 2010	152,793	852,284	385,288	1,079,967	1,279,630	383,305	337,717
Federal Files	93.3	92.7	90.9	91.4	94.8	95.4	95.4
Previous Census Records	0.9	44.0	76.7	73.3	83.8	86.8	87.7
Numident	93.3	92.7	90.9	91.4	94.8	95.4	95.3
HUDCHUMS	0.0	0.0	1.1	5.7	2.5	2.2	2.1
HUDPIC	2.4	3.4	2.2	1.5	1.3	1.3	1.5
HUDTRACS	0.9	0.8	0.7	0.4	0.4	1.1	1.9
IHS	0.8	1.1	1.2	1.1	0.8	0.7	0.5
MEDB	0.0	0.0	0.5	2.3	8.6	92.6	94.3
SSR	0.7	1.6	2.0	1.8	2.7	3.2	3.7
SSS	0.1	0.1	39.3	2.5	0.0	0.0	0.0
TANF	2.5	1.6	0.8	0.5	0.2	0.1	0.0
Third Party Files	0.4	1.4	49.8	76.8	89.5	91.1	89.4
Third Party File 1	0.1	0.0	9.1	35.1	51.8	56.5	53.8
Third Party File 2	0.3	0.8	30.0	51.7	73.6	77.3	73.0
Third Party File 3	0.1	0.5	28.4	49.3	67.4	72.4	69.8
Third Party File 4	0.1	0.2	13.2	25.6	26.9	28.3	27.4
Third Party File 5	0.1	0.1	1.5	52.8	81.3	83.2	78.4
Third Party File 6	-	0.0	0.2	0.6	0.4	0.1	0.0
Third Party File 7	0.1	0.0	4.8	29.6	54.4	63.4	63.4
Third Party File 8	0.1	0.1	1.0	49.3	78.0	78.7	72.1

Sources: Unweighted 2010 ACS and Administrative Records Data

Note: A dash "-" indicates that the cell is suppressed for disclosure avoidance purposes

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

IHS: Indian Health Services

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

SSR: Supplemental Security Income Record

SSS: Selective Service System Registration File

TANF: Temporary Assistance for Needy Families file

Table 9. Administrative Records Coverage of 2010 ACS Sex Data by Administrative Records Source File

Administrative Records Source File	Male		Female	
	Number	Percent	Number	Percent
ACS 2010	2,180,441	100.0	2,290,543	100.0
Federal Files	2,017,996	92.5	2,132,818	93.1
Previous Census Records	1,528,034	70.1	1,659,045	72.4
Numident	2,017,591	92.5	2,132,621	93.1
HUDCHUMS	84,239	3.9	84,454	3.7
HUDPIC	31,530	1.4	51,926	2.3
HUDTRACS	10,493	0.5	20,105	0.9
IHS	20,035	0.9	21,609	0.9
MEDB	362,837	16.6	448,465	19.6
SSS	178,843	8.2	638	0.0
TANF	10,878	0.5	17,731	0.8
Third Party Files	1,431,849	65.7	1,591,083	69.5
Third Party File 1	990,582	45.4	1,102,928	48.2
Third Party File 2	1,238,391	56.8	1,386,602	60.5
Third Party File 3	1,045,765	48.0	1,170,719	51.1
Third Party File 4	441,478	20.2	532,996	23.3
Third Party File 5	1,197,182	54.9	1,335,687	58.3
Third Party File 6	5,403	0.2	6,673	0.3
Third Party File 7	830,101	38.1	906,548	39.6

Sources: Unweighted 2010 ACS and Administrative Records Data

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

IHS: Indian Health Services

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

SSS: Selective Service System Registration File

TANF: Temporary Assistance for Needy Families file

Table 10. Administrative Records Coverage of 2010 ACS Demographic Data for Cases with Allocated ACS Demographic Data

Demographic Characteristics	2010 ACS Persons with Allocated Demographic Response		Coverage by Administrative Records of Allocated Response	
	Number	Percent	Number	Percent
Hispanic or Latino Origin	96,651	2.2	64,481	67.1
Race	76,095	1.7	42,211	55.5
Age	42,304	0.9	20,590	48.7
Sex	6,501	0.1	4,512	69.4

Sources: Unweighted 2010 ACS and Administrative Records Data

Table 11. Agreement between 2010 ACS and Administrative Records Hispanic Origin Response by Source File

Administrative Records Source File	Hispanic			Non-Hispanic		
	Total	In Agreement		Total	In Agreement	
		Number	Percent		Number	Percent
Federal Files						
Previous Census Records	255,227	233,699	91.6	2,813,261	2,798,798	99.5
Numident	N/A	N/A	N/A	N/A	N/A	N/A
HUDCHUMS	20,914	17,147	82.0	140,106	138,076	98.6
HUDPIC	14,441	12,358	85.6	65,087	63,836	98.1
HUDTRACS	253	197	77.9	1,743	1,716	98.5
MEDB	N/A	N/A	N/A	N/A	N/A	N/A
TANF	3,436	2,366	68.9	24,093	23,671	98.2
Third Party Files						
Third Party File 1	134,260	109,821	81.8	1,686,875	1,659,284	98.4
Third Party File 2	197,250	152,612	77.4	2,298,440	2,252,883	98.0
Third Party File 3	151,021	121,045	80.2	1,949,910	1,912,638	98.1
Third Party File 4	72,667	58,208	80.1	833,216	816,607	98.0

Sources: Unweighted 2010 ACS and Administrative Records Data

Notes: N/A is shown in this table for the Numident and MEDB because these files do not have a non-Hispanic category and everyone in these files with a non-missing origin is Hispanic (as described in detail in the data section). Our agreement calculations in the table use a denominator that includes only those with non-missing Hispanic origin. Thus, if we compare individuals who reported Hispanic in the 2010 ACS and individuals coded as Hispanic in the Numident and MEDB, the result would be 100 percent agreement. If we were to calculate agreement rates including everybody (i.e., those with missing Hispanic origin as well as those coded as Hispanic), then agreement rates would be 52.9 percent for the Numident and 26.9 percent for MEDB. For non-Hispanics, we show agreement as N/A because these files do not include a non-Hispanic category, though there are individuals who reported non-Hispanic in the 2010 ACS that are present in these files.

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

TANF: Temporary Assistance for Needy Families file

Table 12. Agreement between 2010 ACS and Administrative Records Race Response by Source File

Administrative Records Source File	Percent of Responses in Agreement by Race						
	White Alone	Black Alone	American Indian or Alaska Native Alone	Asian Alone	Native Hawaiian or Other Pacific Islander Alone	Some Other Race Alone	Two or More Races
Federal Files							
Previous Census Records	96.9	96.1	68.6	94.3	59.5	53.6	35.0
Numident	99.1	98.0	53.6	82.9	72.9	17.2	N/A
HUDCHUMS	98.3	88.7	30.4	70.3	33.6	N/A	2.5
HUDPIC	96.9	96.3	46.9	92.0	60.3	N/A	7.5
HUDTRACS	96.4	95.1	45.9	87.2	-	15.7	9.1
IHS	N/A	N/A	97.9	N/A	N/A	N/A	N/A
MEDB	99.0	97.7	57.2	55.7	42.0	12.3	N/A
TANF	97.2	95.4	79.5	81.8	69.9	N/A	14.4
Third Party Files							
Third Party File 1	98.0	45.0	N/A	85.0	13.5	2.5	N/A
Third Party File 2	97.8	38.8	7.9	73.9	18.2	1.3	N/A
Third Party File 3	95.1	63.2	16.4	80.0	15.4	3.2	N/A
Third Party File 4	94.9	59.5	10.7	80.2	13.0	3.3	N/A

Sources: Unweighted 2010 ACS and Administrative Records Data

Notes: A dash "-" indicates that the cell is suppressed for disclosure avoidance purposes.

N/A indicates that data for a specific group are not available from a particular administrative records source file. For example, several administrative records files do not allow individuals to select more than one race. While there may be individuals who reported Two or More Races in the 2010 ACS that are present in these files, we do not show agreement rates for Two or More Races since that race category is not an option in these files.

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

IHS: Indian Health Services

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

TANF: Temporary Assistance for Needy Families file

Table 13. Agreement between 2010 ACS and Administrative Records Date of Birth Response by Source File

Administrative Records Source File	Date of Birth: Percent in Agreement by Age Group						
	0 to 2	3 to 17	18 to 24	25 to 44	45 to 64	65 to 74	75 and older
Federal Files							
Previous Census Records	83.6	92.1	92.0	92.8	93.8	94.2	92.7
Numident	95.2	95.6	95.2	95.8	95.9	96.5	92.7
HUDPIC	91.3	93.1	92.9	93.9	93.3	91.7	89.4
HUDTRAC	90.9	92.0	93.1	93.2	92.9	93.5	91.2
IHS	92.1	93.6	93.4	94.4	93.7	93.3	91.2
MEDB	-	94.1	94.2	95.5	95.7	96.8	94.7
SSR	94.6	94.8	94.3	93.7	91.7	87.1	82.5
SSS	N/A	N/A	95.5	94.8	N/A	N/A	N/A
TANF	92.4	93.2	93.9	95.5	94.5	91.0	88.3
Third Party Files							
Third Party File 1	-	12.1	76.0	77.0	77.1	76.8	75.8
Third Party File 2	0.0	23.5	63.4	61.5	65.3	69.4	70.3
Third Party File 3	N/A	N/A	56.2	77.2	78.8	79.5	77.9
Third Party File 4	N/A	N/A	61.8	79.9	80.7	81.6	79.3
Third Party File 6	-	-	57.9	58.9	55.1	52.5	67.1
Third Party File 8	-	40.2	82.5	40.2	46.2	52.1	56.3

Sources: Unweighted 2010 ACS and Administrative Records Data

Notes: A dash "-" indicates that the cell is suppressed for disclosure avoidance purposes.

N/A indicates that data for a specific group are not available from a particular administrative records source file. For example, the Selective Service System Registration File (SSS) only contains data on individuals between the ages of 18 and 25. While there may be individuals in this file who reported an age outside this range in the 2010 ACS, we do not show agreement rates for these individuals.

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

IHS: Indian Health Services

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

SSR: Supplemental Security Income Record

SSS: Selective Service System Registration File

TANF: Temporary Assistance for Needy Families file

Table 14. Agreement between 2010 ACS and Administrative Records Sex Response by Source File

Administrative Records Source File	Male			Female		
	Total	In Agreement		Total	In Agreement	
		Match	Percent		Match	Percent
Federal Files						
Previous Census Records	1,515,388	1,508,638	99.6	1,645,841	1,639,085	99.6
Numident	1,998,649	1,988,386	99.5	2,113,890	2,103,608	99.5
HUDCHUMS	83,804	82,675	98.7	83,941	82,261	98.0
HUDPIC	31,186	30,520	97.9	51,545	51,024	99.0
HUDTRACS	10,361	10,181	98.3	19,912	19,744	99.2
IHS	19,909	19,795	99.4	21,498	21,389	99.5
MEDB	360,297	359,163	99.7	444,910	443,735	99.7
SSS	177,084	177,084	100.0	N/A	N/A	N/A
TANF	10,767	10,549	98.0	17,593	17,405	98.9
Third Party Files						
Third Party File 1	984,290	954,908	97.0	1,095,211	1,083,359	98.9
Third Party File 2	1,230,060	1,212,306	98.6	1,376,734	1,357,629	98.6
Third Party File 3	1,038,870	1,013,545	97.6	1,162,500	1,130,680	97.3
Third Party File 4	438,505	427,504	97.5	529,266	514,907	97.3
Third Party File 5	1,189,493	1,165,153	98.0	1,326,455	1,291,652	97.4
Third Party File 6	5,359	5,083	94.8	6,628	6,343	95.7
Third Party File 7	825,016	811,909	98.4	900,274	876,682	97.4

Sources: Unweighted 2010 ACS and Administrative Records Data

HUDCHUMS: Computerized Homes Underwriting Management System file from the Department of Housing and Urban Development (HUD)

HUDPIC: HUD data on persons participating in the public housing program or other rental assistance

HUDTRACS: HUD data on persons receiving rental assistance and participating in other assisted housing programs

IHS: Indian Health Services

MEDB: Center for Medicare and Medicaid's Medicare Enrollment Database

SSS: Selective Service System Registration File

TANF: Temporary Assistance for Needy Families file