

Assessing Availability and Quality of Administrative Records for Asian Americans and Native Hawaiians and Pacific Islanders: Introduction and Federal Database Analysis



Assessing Availability and Quality of Administrative Records for Asian Americans and Native Hawaiians and Pacific Islanders: Introduction and Federal Database Analysis

I. Executive Summary

The role of administrative records—defined as records collected by state, local, or federal government agencies for the implementation of programs—has grown substantially over the last several decades in the planning, implementation, processing, and quality measurement of the United States Census Bureau’s decennial census. Given this rising reliance on administrative records, and the Census Bureau’s current research into even deeper use of them in the 2030 Census, it is important that we understand the strengths and limitations of increased administrative records usage. This research project sets out to understand one key component of these strengths and limitations—the coverage of race information, specifically for the Asian American and Native Hawaiian and Pacific Islander (NHPI) communities, in administrative records.

In order to achieve this goal, we assess the coverage and suitability of federal and state data collections that identify Asian American and NHPI populations. Specifically, we are interested in federal and state administrative records, and the degree to which they not only *include* Asian American and NHPI communities, *but also to what extent they provide disaggregated race data on these populations*. To better evaluate this topic, we focus on governmental census, survey, and administrative records that collect information on individuals.

Extensive research reviewing the effectiveness of survey and census collections for capturing race and ethnicity information exists. Furthermore, the Census Bureau has completed several studies on the use of specific administrative records to help in the enumeration, or the counting of the whole population. Some research has been undertaken on the presence of race and ethnicity data and the coverage of those data in administrative records. To our knowledge, an exhaustive analysis of potential administrative data sources with a specific focus on what racial characteristics are captured does not exist. To fill this gap and accomplish this research, we will develop a robust list of information collected in the federal sector and selected states. We will analyze this list to determine information collections that could enhance the universe of administrative record sources used by the Census Bureau to support the measurement of the total population and improve race and ethnicity coverage, with a focus on Asian Americans and NHPs.

This is the first in a series of white papers that will culminate in a final report. The publications in this series are:

1. Introduction, methodology, and analysis of federal data collections;
2. Analysis of administrative records collected by selected states; and,
3. Final report, with recommendations and conclusions based on the previous analysis.

Ultimately, we seek to answer the question of how valuable these data sources could be in improving the coverage and race classification of Asian American and NHPI populations in the U.S. Census Bureau’s censuses and surveys. As we will discuss in the final installment of this series, prioritizing and focusing on self-response *first* in the decennial census, and augmenting with the linkage of survey and census data to administrative records, may be the most inclusive approach for obtaining the best measure of the total U.S. population and its characteristics. This approach includes:

- Creating a comprehensive address list through administrative data that could be used to contact respondents, thereby improving the coverage of the entire population;
- Conducting additional research on the use of administrative records; and
- Continuing with a focused, self-response first approach to the decennial census, while potentially broadening the usage of linked administrative records to help plan and field the census, and to decrease respondent burden. Linked administrative records combined with a self-response first approach can also

help with processing and imputation, in service of leading to the most fair, accurate, and equitable outcomes possible.

II. Introduction

The increased usage of administrative records in the decennial census raises both opportunities and concerns. It is certainly true that the increased use of administrative records can decrease costs and improve some parts of decennial census operations. However, this is dependent upon which records are used, as well as how, when, and where these records are used. An overreliance on administrative records could lead to inequities in the census based on differential coverage in these records, likely exacerbating historic and persistent differential inaccuracies in the census for communities of color. For instance, a heavy reliance on IRS tax data, without other data sources to augment it, could miss those who are less likely to file taxes, such as low-income households who may not make enough to need to file taxes and who already tend to be harder to count during a decennial census. It is therefore crucial to understand the coverage of different populations in these records, as well as the information that is available about these populations in the same records. As we explain through this analysis, the Bureau’s potential overreliance on administrative records, especially to the detriment of prioritizing self-response, could hinder our ability to accurately count Asian American and NHPI communities. Nonetheless, the increased usage of administrative records, together with an approach that prioritizes self-responses to the decennial census, could improve quality—especially as many states begin collecting disaggregated race data in their own administrative record collections. Moreover, the federal government can, and sometimes does, collect disaggregated data in their collections.

In order to better understand these issues, the data sources included in this review encompass both administrative records and statistical surveys and censuses. Our primary focus is understanding the role of administrative records in augmenting the coverage of censuses and surveys and assessing the potential effects, both positive and negative, of their use on the coverage of Asian American and NHPI communities. By coverage, we mean both that the communities are included in each of these records and that their identities are adequately captured. For example, Asian Americans and NHPs cannot be collapsed into “Other,” and the data needs to be disaggregated by subgroup. This will allow us to better understand if and how administrative records can be used to improve the enumeration of these groups.

Administrative Records, Surveys, and Censuses Defined

Administrative Records (as defined in this research) are records collected by state and federal agencies for the implementation of government programs. These records serve a wide variety of purposes, including, but not limited to, applications for employment and contracts, applications for benefits and services, customer service surveys, licensing applications, and payments of taxes. These data may then be used for a secondary purpose, such as decennial census planning, imputation (the modeling of missing data), or enumeration.

Strengths: These records generally represent a direct interaction with the person filling out a form. Individuals are often informed that the information they provide is controlled by law under penalty of perjury. Furthermore, incorrect submission of information may negatively impact the benefits and services for which the individual is applying.

Weaknesses: What is contained within any single data collection is defined by agency requirements for a specific program. It is not intended to meet a specific statistical data quality standard or to cover an entire population. Furthermore, the population covered in an administrative data set is defined by the coverage of the program or by agency need. It is not meant to be a dataset that enumerates the whole population, meaning that finding full and equitable coverage in these data sources can be an issue. Additionally, a user of this information may not be aware which fields an agency pays greater attention to in collecting quality data; not all fields have equal weight to the agency collecting the data. Finally, users are not always told whether or how a response has been edited.

Administrative Records, Surveys, and Censuses Defined

Surveys and Censuses (as defined in this research) are formal statistical data collections conducted by a government agency through the administration of a questionnaire to a respondent through multiple potential modes such as paper questionnaires, telephone interviews, personal interviews, and internet data collections.

Strengths: What is contained within any single data collection is defined by the requirements of a sponsoring government agency and is intended to meet a specific statistical data quality standard and cover an entire population. A user of this information understands how, and how many, responses have been edited or imputed.

Weaknesses: First and foremost, there is a persistent and differential undercount in both censuses and surveys. Furthermore, these records represent a direct interaction only with the person answering the questionnaire and—excepting the decennial census—there is no penalty for non-response. Often the respondent provides information for their entire household. As a last resort, the Census Bureau may collect information through neighbors, apartment managers, and other individuals in a census. Given the average U.S. household size in 2020 was 2.53 persons, the information for more than 60 percent of individuals in a census process are likely provided by someone else and do not represent self-identification.

In this series of papers, our intent is to add to the conversation on the best approach to accurately capture current population statistics with detailed racial and ethnic characteristics. Specifically, we ask: how and to what extent can administrative records supplement census enumeration and survey responses? A key issue in race and ethnic statistics is self-identification, which occurs through direct collection of information. Definitions of race and ethnicity change over time, as do peoples' views of their own identities. Therefore, when developing current measures of the population, it is always best to acquire the latest available data through surveys and censuses, with up-to-date administrative data used to supplement these data. Given the complementary strengths and weaknesses of these different data sources, combining multiple-source administrative records with current and prior survey and/or census data may present the best approach to a fair and accurate measurement of the total population and its characteristics. However, this approach is only possible if the requisite data are available in administrative records for this purpose.

To better understand the overall coverage of Asian Americans and NHPs in these products, we begin by discussing the collection of race and ethnicity information by the Census Bureau. We review prior research on administrative records by the Census Bureau and others. We then discuss the methodology we employ to better understand the potential coverage of Asian Americans and NHPs in federal and state administrative data collections. Finally, we share our analysis of Asian American and NHP coverage in federal data collections.

III. History and Prior Research on the Census Bureau’s Measurement of Race and Use of Administrative Data

A. Census Bureau’s Measurement of Race and Ethnicity

The Census Bureau has collected information from individuals on race since the inception of the decennial census in 1790.¹ Beginning in 19th century, the Census Bureau began collecting information on what we now identify as the Asian American population. The Census first collected information on the Chinese population in 1860—and only in California. The growing number of Chinese immigrants who came to the United States in the 1870s and 1880s prompted the Census to expand the category beyond California to include the entire country. In 1890, an additional national origin group, Japanese, was added. An uptick in the number of Korean, Filipino, and Asian Indians recorded in the “Other” category in 1910 promoted their inclusion in the 1920 Census. However, the term “Hindu” was used instead of Asian Indian—which the Census notes is the “only time that a religious term has been included as a race question category in a U.S. decennial census.”² By 1950, the Census had removed Korean and Hindu.

In 1960, after Hawaii and Alaska both became states, data were first collected on the Native Hawaiians and Pacific Islander population, but only in Hawaii. This question was used throughout the United States in 1970 forward. With increased Korean immigration, Korean was added again in 1970. The term Asian American was coined by activists in 1968,³ and this may have pushed the Office of Management and Budget (OMB) to issue Statistical Policy Directive No. 15 (SPD 15) in 1977 to change the race category to “Asian or Pacific Islander” for the 1980 Census.⁴ It also included options to select Vietnamese, Guamanian, and Samoan. In 1990, the Census included a write in option for “Other API” and included Hmong, Fijian, Laotian, Thai, Tongan, Pakistani, and Cambodian as examples. When OMB revised SPD 15 in 1997—the last time the standards were updated—“Asian” and “Native Hawaiian or Other Pacific Islander” were separated.

Since 1970, the Census Bureau has also collected information on ethnicity—i.e., people who self-identify as Hispanic or Latino. These race and ethnicity categories have changed significantly over time,⁵ and current Census Bureau classifications remain consistent with the minimum race categories mandated by OMB in 1997.⁶ Standard tabulations currently available in the 2020 Census include ethnicity (Hispanic/non-Hispanic) cross-tabulated by race, including the ability to select multiple of the following races:⁷

¹ https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html

² Ibid.

³ <https://www.nbcnews.com/news/asian-america/after-50-years-asian-american-advocates-say-term-more-essential-n875601>

⁴ The Race and Ethnic Standards for Federal Statistics and Administrative Reporting, known more commonly as Statistical Policy Directive No. 15, represented the first time the federal government formally created an Asian race category for minimum reporting. The directive was subsequently revised in 1997 and is currently under review by the Office of Management and Budget.

⁵ Scholarship on U.S. census history that discusses racial formation includes Anderson, Margo J. *The American Census: A Social History*. Second edition. New Haven: Yale University Press, 2015; Bouk, Dan. *Democracy’s Data: The Hidden Stories in the U.S. Census and How to Read Them*. First edition. New York: MCD, 2022; Hochschild, Jennifer L., and Brenna Marea Powell. “Racial Reorganization and the United States Census 1850–1930: Mulattoes, Half-Breeds, Mixed Parentage, Hindoos, and the Mexican Race.” *Studies in American Political Development* 22, no. 1 (2008): 59–96; Mora, G. Cristina. *Making Hispanics: How Activists, Bureaucrats, and Media Constructed a New American*. Chicago: The University of Chicago Press, 2014; Nobles, Melissa. *Shades of Citizenship: Race and the Census in Modern Politics*. Stanford, CA: Stanford University Press, 2000; Prewitt, Kenneth. *What Is “Your” Race?: The Census and Our Flawed Efforts to Classify Americans*. Princeton, NJ: Princeton University Press, 2013; “The Census Counts, the Census Classifies,” in Foner, Nancy, and George M. Fredrickson, eds. *Not Just Black and White: Historical and Contemporary Perspectives on Immigration, Race, and Ethnicity in the United States*. New York: Russell Sage Foundation, 2004; Schor, Paul. *Counting Americans: How the US Census Classified the Nation*. New York, NY: Oxford University Press, 2017; Thompson, Debra. *The Schematic State: Race, Transnationalism, and the Politics of the Census*. Cambridge, United Kingdom: Cambridge University Press, 2016.

⁶ <https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf>

⁷ <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>

- White
- Black or African American
- American Indian or Alaskan Native
- Asian
- Native Hawaiian and Pacific Islander
- Some Other Race

On September 23, 2023, the Census Bureau released detailed race and ethnicity information from the 2020 Census for 270 detailed race groups, 30 detailed ethnic groups, and 1,187 American Indian and Alaskan Native tribes and villages.⁸ Of those detailed race groups, 41 were Asian groups and 31 were Native Hawaiian and Pacific Islander groups. By comparison, the 1980 Census, which was the first census to implement the Race and Ethnic Standards for Federal Statistics and Administrative Reporting released in 1977, reported 12 Asian groups and 7 Native Hawaiian and Pacific Islander groups. These detailed race categories are often defined by national origin groups. Statistics for detailed race and ethnicity groups are also provided in the American Community Survey 1-year and 5-year products.⁹ They are included to a lesser extent in other surveys such as the Current Population Survey Annual Social and Economic Supplement¹⁰ and the Survey of Income and Program Participation.¹¹

A. Looking Towards the 2030 Census: Administrative Records

Given that the decennial census is our country's best point-in-time measure of our population, it is important that the process is as effective as possible at obtaining fair and accurate data on race and ethnicity. The use of administrative data has the potential to enhance the decennial census enumeration process in several ways if used appropriately.¹² The first step could be to create a comprehensive address list through administrative data such as: the USPS Delivery Sequence File;¹³ federal, state, and potentially privately held records; the Local Update of Census Address operation, the New Construction Program, Imagery, and field operations.¹⁴ This approach could provide the Census Bureau with the best chance of contacting respondents, thereby improving coverage of the entire population. Additionally, administrative records contain much more information than just addresses. They include identifying information such as the linkage between name, address, and telephone numbers in a temporal context. They often have some information on a person's age, sex, race, and ethnic characteristics. However, unless this information is coming from a previous census or survey, it likely will not be disaggregated by different racial subgroups. In the 2020 Census, the Census Bureau employed selected administrative data to define addresses and limited their use in providing person-level coverage and characteristics to the Non-Response Follow-Up operation in a very targeted manner.

As part of our nation's democratic process, it is essential that individuals can represent themselves and their characteristics. Their statistics impact the apportionment of political power,¹⁵ voting rights,¹⁶ and the distribution of

⁸ <https://www.census.gov/library/stories/2023/09/2020-census-dhc-a-race-overview.html>

⁹ https://usa.ipums.org/usa-action/variables/RACE#codes_section

¹⁰ <https://www2.census.gov/programs-surveys/cps/techdocs/questionnaires/Demographics.pdf>

¹¹ https://www2.census.gov/programs-surveys/sipp/questionnaires/2022/2022_SIPP_Instrument_Specifications.pdf

¹² There are also important improvements to the enumeration process that rely on other approaches, such as improved community engagement. While we do not discuss these improvements here, they are both necessary and complementary to the application of administrative data.

¹³ <https://firstlogic.com/insights/a-business-guide-to-usps-dfs2-processing>

¹⁴ <https://www.census.gov/newsroom/press-kits/2018/2020-adcan.html#:~:text=Address%20Canvassing%20is%20Now%20Conducted,Census%20Bureau's%20Master%20Address%20File.>

¹⁵ <https://www.census.gov/topics/public-sector/congressional-apportionment.html#:~:text=The%20Constitutional%20basis%20for%20conducting,Representatives%20among%20the%2050%20states.>

¹⁶ <https://www.archives.gov/milestone-documents/voting-rights-act>

over \$2.8 trillion in federal dollars annually,¹⁷ in addition to the allocation of state and local resources. Therefore, the Census Bureau must be able to contact those individuals and collect their demographic characteristics through censuses and surveys.

However, census and survey processes are not perfect. Some reasons people may be undercounted are because they are not contacted or refuse to respond, responses are provided by proxies, or incomplete or incorrect responses are provided.¹⁸ Federal and state administrative records are interactions between individuals and the governments representing them. They are self-representing and are part of our nation’s historical record—owned by the people and for the people—in operations that are monetarily supported through personal fees and taxation. In the conducting of a census or survey operation, the Census Bureau assumes they do not know any information about an individual upon contact. In reality, information exists from prior decennial census and survey responses, and administrative data can and have helped in census or survey operations. The use of this preexisting information can also potentially be expanded. But first, we must know what data do—and do not—appear in administrative records. The Census Bureau may then be able to compare that information to survey and census collections to potentially improve the efficiency, cost, fairness, and accuracy of final statistics. This can be achieved by improving both direct data collection activities as well as the processing of administrative data.

B. Census Bureau Administrative Records: The “Early Days”

The Census Bureau began developing its administrative records infrastructure in the mid-1970s with the Standard Statistical Establishment List (SSEL). Now called the Longitudinal Business Database, this list consists of both employer and non-employer U.S. businesses.¹⁹ The use of demographic administrative records took a slightly different approach. Individual records were first received from the Internal Revenue Service (IRS) in the form of personal tax return data (IRS 1040 Form). This supported the Population Estimates Program’s need to produce population and per capita income estimates for substate political entities as part of federal general revenue sharing.²⁰

Development of an administrative records infrastructure for individuals that was similar to the SSEL for businesses did not begin until the mid-1990s. In 1994, the Census Bureau founded the Team for Administrative Records Planning. The following year, their efforts culminated in the creation of the Administrative Records Research Staff. This staff was tasked with the research and development of administrative records for testing in the 2000 Census. This staff conducted seminal research in the mid- to late-1990s on the potential use of numerous national-level files from federal agencies. They received, processed, and analyzed almost two billion records from sources including:

- Internal Revenue Service personal tax and selected information returns files;
- Social Security Administration (SSA) Numident file;
- Housing and Urban Development (HUD) Tenant Rental Assistance Certification System;
- Centers for Medicare and Medicaid Services (previously known as the Health Care Financing Administration) Medicare beneficiaries database;
- Indian Health Service registration file; and
- Selective Service System registrant file.

¹⁷ <https://www.census.gov/library/working-papers/2023/dec/census-data-federal-funds.html>

¹⁸ <https://nap.nationalacademies.org/catalog/27150/assessing-the-2020-census-final-report>

¹⁹ <https://www2.census.gov/ces/wp/2021/CES-WP-21-08.pdf>

²⁰ <https://www.govinfo.gov/content/pkg/STATUTE-86/pdf/STATUTE-86-Pg919.pdf>, and <https://www.finance.senate.gov/imo/media/doc/ConfRpt92-1229.pdf>

The Census Bureau developed the Statistical Administrative Records System (STARS) and performed an Administrative Records Experiment in 2000²¹ on two test sites—two counties in Maryland and three counties in Colorado. In this experiment, administrative records were compared to 2000 Census responses in the two test sites with mixed results. County-level total population counts were consistent with decennial results. However, results for geographic areas below the county level were less successful. Seventy percent of census tracts and less than 40 percent of census blocks were within 5 percent of the decennial census population counts. These numbers can be explained, in part, because the records were only accepted if they matched the Master Address File.²² Also, children were undercounted in administrative records, and lacked any direct race or ethnicity reporting.²³ Results by race and ethnicity were different from the decennial census. The stated reason for these differences was that the imputation methodologies used to assign race to administrative records at that time were deficient, especially for children.

Expanded research continued in the first decade of the 2000s²⁴ with the addition of the Department of Education’s Free Application for Student Aid file, HUD’s Computerized Underwriting Mortgage file, the Medicaid Statistical Information System, and the United States Postal Service’s National Change of Address file.²⁵

The Person Validation System (PVS) is the backbone of the Census Bureau’s ability to link administrative, survey, and census response data. First developed in the late 1990s, it has been continually enhanced since and was a key component of the Census Bureau’s development and use of administrative records.²⁶ The PVS has two major features. The first is the verification of person-based records against the SSA’s Numident file through probabilistic matching using name, address, and date of birth. The second is the removal and replacement of personal identifiers with a unique Protected Identification Key (PIK) for each person.

C. Census Bureau Administrative Records: 2010 Match Study

The next major milestone in Census Bureau administrative records research was the completion of a national-level study in 2012 that matched administrative records to 2010 Census results.²⁷ Key components of assessing administrative records coverage are the ability to develop an unduplicated roster of people, the ability to geolocate those people, and finally the ability to attribute demographic characteristics to those people.

The 2010 Match Study showed that administrative records performed well at identifying decennial census addresses and included approximately 20 million more addresses than the actual decennial census. These addresses may not have matched due to the inclusion of P.O. boxes (which are not valid for purposes of census questionnaire delivery), new construction not included in the decennial address database, and different descriptions for the same address that did not match. Administrative records sources included prior decennial census records, the STARS system, and third-party (commercial) data. Finally, administrative records data matched 92.6 percent of 2010 Census addresses and 88.6 percent of persons. Administrative records match rates to the 2010 Census were similar for both Hispanic (94.2 percent) and non-Hispanic householders (94.7 percent). Householders who reported their race as Asian alone matched at a rate of 95.8 percent and the match rate for NHPI alone was 93.5 percent.²⁸ In looking at person-level data, 11.6 percent of

²¹ <https://www2.census.gov/programs-surveys/decennial/2000/program-management/5-review/txe-program/2000-arex-admin-records-outcomes.pdf>

²² The Master Address File is the Census Bureau’s key address list for undertaking the decennial census and building address frames for other surveys.

²³ <https://www2.census.gov/programs-surveys/decennial/2000/program-management/5-review/txe-program/2000-arex-admin-records-outcomes.pdf>

²⁴ <https://ncvhs.hhs.gov/wp-content/uploads/2014/05/060918p1.pdf>

²⁵ Much of the work conducted during this period was documented in internal memoranda and currently does not appear to be publicly available. One of this paper’s authors was involved extensively in administrative records research, and in the design and creation of the infrastructure called the Census Bureau’s Statistical Administrative Records System.

²⁶ <https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf>

²⁷ https://www.census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_247.pdf

²⁸ These match rates are for the broad race categories of Asian American and NHPI; the study is silent on matches for detailed race.

records categorized as Asian, and 16.1 percent categorized as NHPI were unable to be matched to the 2010 Census because they lacked identifiable information to assign them a PIK. In contrast, 7.6 percent of records characterized as white alone could not be matched.

D. Census Bureau’s Internal Kid-Link Database

One significant improvement in the attribution of race information to administrative records used in the postcensal population estimates (as well as, to a lesser degree, the Demographic Analysis measures) is the Census Bureau’s internal database Kid-Link. This dataset links persons under the age of 17 with their parents living in the same household to assign race and ethnicity.²⁹ In the early stages of the Census Bureau’s administrative records development, SSA’s Numident file was a major component of the attribution of race information to administrative records, even though its race information was limited and categories changed over time. Furthermore, in 1987, SSA eliminated race information collection for children with its Enumeration at Birth system.³⁰ The basic methodology of this file is to use administrative record-matching between SSA’s Numident file and other administrative record files combined with decennial census records. Race and ethnic characteristics are then modeled for children on birth records based on their parents’ race(s).³¹

This work at the Census Bureau is currently done only for the minimal race categories defined by the 1997 OMB standard, which is limited due to the lack of detailed, disaggregated data. However, as we will see in the state chapter, some states are beginning to collect parents’ race information on birth certificates by disaggregated race groups. This will provide an interesting—but currently uneven—view into disaggregated race data for young children and their parents in *some* states. Depending on which states provide this level of detail and which do not, further inequities may occur.

E. Expansion of Administrative Records Repository: Census 2020

The Census Bureau expanded its administrative record repository in the mid-2010s. Production Environment for Administrative Records Staging, Integration and Storage (PEARSIS) integrated data from a multitude of federal, state, and third-party data sources. The Census Bureau collected these data in the decade leading up to the 2020 Census operations including address lists, verification of enumeration activities, and Non-Response Follow-Up.³² The PEARSIS system is based on an aggregation of 63 billion records containing individuals and addresses. It uses a rule-based approach to choose the correct geographic location for a person based on multiple records with potentially different addresses (e.g., a simple rule could be: “use the most recent record”). The 2020 population coverage³³ achieved by PEARSIS was significant—totaling 331.6 million persons as compared to the decennial census count of 331.4 million persons. Furthermore, the mean net coverage rate for states was 99 percent with a mean absolute percent error of 1.8 percent. This was a significant improvement over prior analysis. County-level statistics did not fare as well—the mean absolute percent error (MAPE) for counties was 5.4 percent. More populous counties performed better (3.4 percent MAPE) than smaller counties (7.4 percent MAPE). Measures of the population by race and ethnicity were not presented.

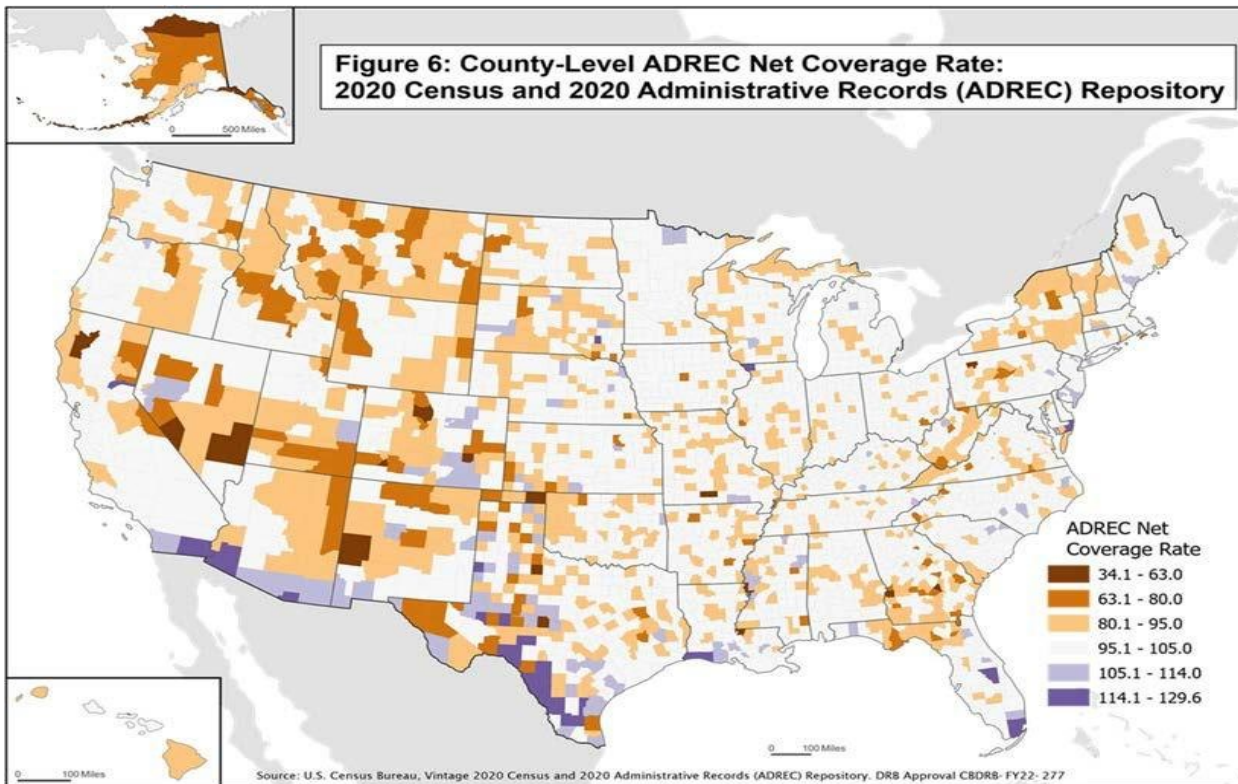
²⁹ <https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-07.pdf>

³⁰ <https://www.ssa.gov/dataexchange/documents/Updated%20State%20Processing%20Guidelines%20for%20EAB.pdf>

³¹ <https://www2.census.gov/news/press-kits/2019/so-demographers-assoc-meeting/presentations/race-of-children-in-population-estimates.pdf>; https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020da_methodology.pdf.

³² Ortman, Jennifer M., Kevin M. Shaw, Rajini L. Kotha, and John L. Boies, “Compiling, Enumerating, and Estimating: Evaluating the 2020 Administrative Records Repository.” Joint Statistical Meetings, 2022, Washington, DC.

³³ Population coverage was measured as the net coverage rate—or the difference between PEARSIS and the decennial census divided by the decennial census times 100.



Source: United States Census Bureau Map, presented at the Joint Statistical Meetings, 2022.³⁴

F. Real-Time 2020 Administrative Record Census Simulation

The Real-Time 2020 Administrative Record Census Simulation³⁵ was developed to continue the exploration of how administrative records can help the Census Bureau improve coverage in the decennial census. It uses similar data to PEARSIS but employs a probabilistic approach to the assignment of individuals to particular geographic locations. For example, this approach would take all records for a person and build a statistical model that says what the most likely address of the person would be. Those geographic locations are limited to addresses included in the Census Bureau’s Master Address File. In addition, there are record sources for individuals without Social Security or Taxpayer Identification Numbers. These include “USCIS [U.S. Citizenship and Immigration Services] naturalizations and lawful permanent residents, ADIS [Arrival and Departure Information System], State Department passports and Worldwide Refugee Admissions Processing System, Bureau of Prisons, Bureau of Justice Statistics National Corrections Reporting Program, U.S. Marshals Service, Department of Interior Incident Management Analysis Reporting System, Medicare Enrollment Database, HUD, Indian Health Services, Selective Service System, Veterans Affairs, and Nebraska driver’s licenses.”³⁶ The inclusion of these records resulted in an administrative records population that was 2.3 percent higher than the 2020 Census. The Asian alone population in the 2020 Census linked directly to administrative records 79 percent of the time and the NHPI alone population linked at a rate of 71 percent, compared to 86 percent for the white alone population. The only population category that performed worse than the NHPI category was the Some Other Race category that linked at a rate of 64 percent. In the estimation process, certain demographic groups were underrepresented. These included persons in the 65–74 age group (470,600 fewer in the Administrative Record census), non-Hispanic Asians (153,400 fewer), and non-Hispanic Two or More Races (400,200 fewer).

G. Conclusions of Prior Administrative Records Research

³⁴ Ibid.

³⁵ <https://www2.census.gov/programs-surveys/decennial/2020/program-management/evaluate-docs/EAE-2020-admin-records-experiment.pdf>

³⁶ Ibid.

The 2010 Match Study showed that limited improvements could be made in both the definitions of households through address attribution and the ability to link records for both Asian American and NHPI populations. The PEARSIS study showed that through the direct use of administrative records, a total population roster is close to development. However, no information was provided on the coverage of race and ethnicity, meaning that it is unclear if the increases in coverage apply equally for all groups. The Real-Time 2020 Census Simulation study—which included additional files—appears to have improved on the PEARSIS results for total population measurement. Results for Asian American and NHPI populations could be improved in total. At the same time, these results could limit our understanding of disaggregated race for those individuals for whom data were not directly collected in the decennial census. The simulation study was designed to be an estimation process. Therefore, it employed extensive modeling and could prove quite useful for updating age, sex, race, and ethnicity estimates as well as survey controls. It is highly questionable if this approach would pass the litmus test of representing individual responses in a decennial census context.

IV. Asian American and NHPI Coverage in Federal Administrative Records

Based on the Census Bureau’s stated hope to use more administrative data in the 2030 Census, along with our current knowledge of the coverage of Asian American and NHPI communities in administrative records, additional research is needed to determine the overall coverage of, as well as the ability to report disaggregated data on, Asian American and NHPI communities. Below we outline why administrative records and federal data collections covered by the Paperwork Reduction Act (PRA) are a great place to start. Focusing on state-level records, which may be harder to find and collate information on, is also a necessary step. Together, assessing both federal and state data will improve our understanding of the coverage of all communities in administrative data collections.

There are numerous data collection forms used by governments for program administration. Not all forms contain personally identifiable information (PII) – in fact, most do not collect it. We will review all federal collection forms and include those that contain PII in our inventory. Also, because we are looking for information that could potentially be used across the nation, our review is limited to those forms with an estimated coverage of 10,000 submissions or greater. We will review each form for who would be filling out the form (i.e., its universe), purpose, expected population of interest, and content. This will comprise our database of information (defined in the Appendix). We will identify forms as being “in-scope” if they contain race, ethnicity, or language information, or “potentially in-scope” if these forms only contain information that allows them to be linked to other data sources.³⁷ Forms that do not meet these conditions will be initially identified as “out-of-scope” for this research. However, all forms will remain in our database.

We will next review the forms defined as in-scope or potentially in-scope to determine additional data items (as outlined in the Appendix) that are useful in this research. We will add further detail from the form where available. The final step is conducting a meta-analysis of this information. We give an overall view of the potential coverage of the Asian American and NHPI populations in the forms that have been reviewed, and cover the level of race data disaggregation that is present. In this paper, we conduct this review on federal government administrative records and report our findings. The next paper will conduct a similar review on state government administrative records in select states and the District of Columbia.

A. Federal Administrative Records Analysis: Methodology

Our goal is to determine the federal data collections that identify Asian American and NHPI populations, as these data sources could ultimately improve the coverage of race (and potentially subgroup) classification of these populations in the U.S. Census Bureau’s censuses and surveys. We conducted an extensive study of potential administrative data sources that may provide the Census Bureau with an enhanced ability to capture the total population race and ethnic characteristics.

The federal government mandates that all data collection processes adhere to the requirements of the Privacy Act as amended,³⁸ as well as the PRA.³⁹ The process necessitates that the collection agency submits their form for review by OMB. This includes collections that are either administrative records or surveys and censuses. After initial review, OMB must evaluate these forms every three years.⁴⁰ The forms and the data they collect are controlled by the Federal

³⁷ While they do not contain race or ethnicity data, the fact that potentially in-scope records have other information (including linkable characteristics such as SSN or name) means that they may *augment* other data sources that do have race and ethnicity data. For instance, a form without race and ethnicity may have a more up to date address that will allow for better geographic representation of the community.

³⁸ <https://epic.org/the-privacy-act-of-1974/>

³⁹ <https://pra.digital.gov/>

⁴⁰ <https://www.whitehouse.gov/omb/information-regulatory-affairs/federal-collection-information/>

Records Act⁴¹ and are part of our nation’s historical record. Access to these records may be controlled by both the National Archives and Records Administration as well as the custodial agency and their confidentiality processes.⁴² OMB reviews a wide variety of data collections conducted by the federal government. The total number of active forms that have been approved as denoted by their control numbers (in active status)⁴³ is in excess of 10,000.⁴⁴ We began by accessing OMB’s Inventory of Approved Data Collections⁴⁵ to select the universe of records for review on August 6, 2023. We chose 985 collections and filed Information Collection Reviews (ICRs) based on the following criteria – the collection:⁴⁶

- Satisfied the definition of PII as defined by OMB Circular No. A-130;
- Included a form that required a Privacy Act Statement [5 U.S.C. §552a(e)(3)];
- Affected individuals or households;
- Anticipated a collection of information from more than 10,000 respondents; and,
- Included Information Collection Requests with assigned control numbers that were active (currently approved by OMB), historically active or inactive (previous reviews that are or are not in the active inventory), and those currently under review marked as received.

We reviewed each collection request, related forms, and supporting materials for their intended population universe, purpose, expected number of respondents, and content. We then created a database of their characteristics based on the findings of our review. All collection requests were retained in our database regardless of their classifications as defined below.

- The first step in our review was to identify whether an information collection request contained usable PII for the purposes of this research. We defined usable PII in a collection request as form(s) containing information that could allow an individual’s information to be linked to other data sources for the purpose of identifying and deduplicating population records both within and across sources. PII could also be used for augmenting other data sources that do contain race information. These forms included information such as name and/or Social Security Number. We initially classified 819 of the 985 ICRs as being “potentially usable” given that the information they contained could be linked to other data sources. The 166 ICRs that did not contain usable PII we classified as “out of scope.”
- The second step was to review information collections defined as “potentially usable.” This would allow us to determine if those forms also contained location information—such as addresses and telephone numbers—to assist in identifying and deduplicating population records. These “potentially usable” records could be linked to other records or information that could assist the Census Bureau in both Master Address File development and respondent contact strategies. We classified this subset of 750 ICRs as “potentially usable and linkable with location.”
- The third step was to review the 818 ICRs that were linkable (with or without location information) to determine if they contained information on race. These were initially classified as “in-scope for race analysis.” There were 142 ICRs that contained questions about a person’s racial characteristics—133 of which contained respondent name or SSN with location information, while 9 contained only name or SSN.

⁴¹ <https://www.archives.gov/news/topics/federal-records-act>

⁴² Some state governments adhere to similar practices as the federal government. Others do not. Therefore, our methods of discovery will vary for the states selected for review.

⁴³ Control numbers are form identification numbers that OMB uses as a way to catalogue forms they have or are reviewing.

⁴⁴ <https://www.reginfo.gov/public/do/PRARReport?operation=11>

⁴⁵ <https://www.reginfo.gov/public/do/PRARReport?operation=11>

⁴⁶ We did not include forms that included national security information from the Department of Defense; these forms would be unlikely to be shared.

The remaining 677 ICRs that did not specifically request racial characteristics contained information that, while out of scope for this analysis, could be useful to improve household information for the population overall—including persons identifying themselves as Asian American or NHPI.

- We then reviewed these 142 ICRs to determine the demographic characteristics (such as age, race, sex, and ethnicity) and details of those characteristics for each collection. The results of this analysis are presented in the following section.

B. Federal Administrative Records Analysis: Findings

1. Agencies Collecting PII

The 985 ICRs we reviewed came from numerous U.S. government agencies with a wide range of purposes for collection. These included statistical surveys, customer satisfaction surveys, program evaluations, and program administration to provide services or regulate activities.⁴⁷ Table 1 provides a summary of ICRs that we reviewed broken down by agency. The vast majority of these data collections requested PII, including identifiers such as name and/or SSN, as well as location and contact information, such as address and/or telephone numbers.

Table 1. Department-Level Summary of Information Collection Requests

Department	Linkable with		Not Linkable	Grand Total
	Linkable	Location		
Social Security Administration	9	114	6	129
Veterans Administration	5	107	12	124
Department of Defense	8	61	28	97
Department of Homeland Security	5	79	11	95
Department of Health and Human Services	9	60	20	89
Department of Treasury	7	55	11	73
Department of Commerce	3	31	9	43
Department of Education	3	33	6	42
Department of State	6	29	2	37
Department of Labor	4	21	11	36
Office of Personnel Management	2	28	4	34
Department of Agriculture	3	22	9	34
Department of Justice	1	18	9	28
Department of Transportation	1	19	5	25
Department of Interior	0	18	3	21
Federal Communications Commission	0	8	6	14
Other Agencies	3	46	13	62
All Departments	69	749	165	983

Note: Departments and agencies with less than 10 Information Collection Requests are combined into "Other Agencies"

2. Assessing the Utility of Federal Information Collections for Improving Asian American and NHPI Statistics

⁴⁷ We removed two of these ICRs from our analysis: the Special Supplemental Nutrition Program for Women, Infants, and Children; and the National Youth in Transition Database, because they referenced state and local level data collections that were not fully defined. These state and local data collections will be analyzed in the next paper in this series.

Many of the federal ICRs include race data. We next determine what data collections might improve coverage for the Asian American and NHPI populations and assess access to this information. For example, acquiring information from the Department of Defense would be extremely difficult, if not impossible, due to national security considerations. Those data collections (whether they included race information or not) were classified as Out of Scope. Additionally, data collections that did not contain race information and were expected to be relatively small (less than 250,000 estimated responses) were classified as out of scope. Table 2 details the number of data collections by their inclusion of race information and the ability to improve rosters (i.e. the listing of who lives at a particular address). Many administrative records have information to improve rosters, which is why administrative data are used by the Census Bureau to build out their address and housing roster frames for the decennial census and American Community Survey. They are also used for the numerous other demographic and housing surveys the Census Bureau fields. However, those records are far less likely to provide disaggregated race information than survey data collections. Given recent events, such as the fire in [Maui](#), one example of a roster improvement that could be valuable for NHPI statistics is FEMA’s collection of disaster assistance registrations⁴⁸.

Table 2. Federal ICR's Classified by Their Potential to Improve Asian Americans and NHPI Rosters and Characteristics

Classification	Administrative Records	Statistical Studies	Total
Race Data and Linkable With Location Information	41	59	100
<i>Disaggregated for Asian Americans and NHPI</i>	1	19	20
<i>1997 OMB Categories</i>	39	38	77
<i>Undefined</i>	1	2	3
Race Data and Linkable Without Location Information	3	6	9
<i>Disaggregated for Asian Americans and NHPI</i>	0	2	2
<i>1997 OMB Categories</i>	3	3	6
<i>Undefined</i>	0	1	1
Lacks Race Data / Usable for Roster Improvement	200	4	204
Out of Scope	600	70	670
<i>Disaggregated for Asian Americans and NHPI</i>	0	1	1
<i>1997 OMB Categories</i>	0	2	2
<i>Lacks Race Data</i>	600	67	667
Grand Total	844	139	983

Note: Out of Scope includes ICR's for National Security Collections, Non-Linkable Records, and records limited to roster improvement with <250,000 estimated respondents

Table 3 describes the estimated number of responses agencies expect to receive as part of their data collections for the subset of ICRs described in Table 2. Responses are defined as completed interactions with a form (paper or internet), survey questionnaire, telephone interview, or personal contact. A response could represent an individual or an individual reporting for every person in their household. The federal government collects an impressive amount of data

⁴⁸ https://www.reginfo.gov/public/do/PRAICList?ref_nbr=201708-1660-004

from forms containing race information (342 million responses). This figure is based on the estimated responses included in their ICRs—120 million administrative records and 222 million survey forms. Additionally, data from over three billion forms from both administrative records and survey data could be used to improve address and contact information for surveys and censuses. The table shows data broken down by three possible types of race reporting: disaggregated data for Asian Americans and NHPI, data reported by the 1997 OMB categories described in the history section above, and undefined race, which means the form has a spot for race on it, but it does not provide check boxes or examples of what responses are valid.

Table 3. Estimated Federal Data Collection Response Forms Classified by Their Potential to Improve Asian Americans and NHPI Rosters and Characteristics (In Millions)

Classification	Administrative		Total
	Records	Statistical Studies	
Race Data and Linkable With Location Information	119.87	216.81	336.67
<i>Disaggregated for Asian Americans and NHPI</i>	0.11	202.95	203.06
<i>1997 OMB Categories</i>	63.33	6.64	69.97
<i>Undefined</i>	56.43	7.21	63.64
Race Data and Linkable Without Location Information	0.23	5.54	5.78
<i>Disaggregated for Asian Americans and NHPI</i>	0.00	4.71	4.71
<i>1997 OMB Categories</i>	0.23	0.82	1.05
<i>Undefined</i>	0.00	0.01	0.01
Lacks Race Data / Usable for Roster Improvement	3,227.07	34.83	3,261.90
Out of Scope	3,281.48	11.86	3,293.34
<i>Disaggregated for Asian Americans and NHPI</i>	0.00	1.30	1.30
<i>1997 OMB Categories</i>	0.00	0.04	0.04
<i>Lacks Race Data</i>	3,281.48	10.52	3,292.00
Grand Total	6,628.65	269.04	6,897.69

Note: Out of Scope includes ICR's for National Security Collections, Non-Linkable Records, and records limited to roster improvement with <250,000 estimated respondents

Table 4 describes which agencies collect race information in their forms. The Census Bureau collects the most information on disaggregated race. However, other agencies collect over five million forms with disaggregated race and over 70 million forms with race categories that are consistent with the 1997 OMB standards.

Table 4. Sources of Linkable Federal Data Collections Classified by Their Potential to Improve Asian American and NHPI Rosters with Characteristics (In 1000's of Estimated Forms)

Department/Subagency	Disaggregated	1997 OMB Categories	Undefined	Grand Total
DHS/Transportation Security Administration	0.0	0.0	7,094.5	7,094.5
DHS/Coast Guard	0.0	160.0	0.0	160.0
DHS/Citizenship and Immigration Services	0.0	7,866.8	0.0	7,866.8
DOC/Census Bureau	202,294.6	1,349.3	0.0	203,643.9
Department of Energy Office of Administration	0.0	15.0	0.0	15.0
DOI/U.S. Fish and Wildlife Service	0.0	220.6	0.0	220.6
DOJ/Alcohol, Tobacco, Firearms, and Explosives	0.0	18,401.1	0.0	18,401.1
DOJ/Office of the Pardon Attorney	0.0	20.0	0.0	20.0
DOJ/Office of Justice Programs	0.0	12.3	0.0	12.3
DOL/Bureau of Labor Statistics	604.5	12.6	0.0	617.1
DOL/Civil Rights Center	0.0	0.0	56,425.5	56,425.5
DOL/Employment and Training Administration	0.0	347.0	0.0	347.0
DOT/Federal Aviation Administration	0.0	180.0	0.0	180.0
DOT/National Highway Traffic Safety Administration	0.0	147.4	0.0	147.4
ED/Institute of Education Sciences	151.0	1,675.3	132.4	1,958.7
Federal Communications Commission	0.0	255.5	0.0	255.5
Federal Reserve System	0.0	25.0	0.0	25.0
HHS/Administration for Children and Families	0.0	3,444.0	0.0	3,444.0
HHS/Centers for Disease Control and Prevention	265.3	814.9	0.0	1,080.2
HHS/Centers for Medicare & Medicaid Services	243.5	799.7	0.0	1,043.1
HHS/Food and Drug Administration	123.4	209.7	0.0	333.1
HHS/Office of the Assistant Secretary for Health	0.0	200.0	0.0	200.0
HHS/Office for Civil Rights	0.0	13.8	0.0	13.8
HUD/Office of Policy Development and Research	88.3	101.9	0.0	190.2
National Science Foundation	0.0	736.7	0.0	736.7
Office of Personnel Management	0.0	4,068.9	0.0	4,068.9
Peace Corps	0.0	23.0	0.0	23.0
Social Security Administration	0.0	14,003.8	0.0	14,003.8
STATE/Administration of Foreign Affairs	0.0	12.0	0.0	12.0
USDA/Office of the Secretary	0.0	1,910.0	0.0	1,910.0
USDA/Food and Nutrition Service	0.0	13,016.1	0.0	13,016.1
USDA/Food Safety and Inspection Service	0.0	73.4	0.0	73.4
Veterans Administration	4,000.0	910.6	0.0	4,910.6
Grand Total	207,770.5	71,026.2	63,652.4	342,449.1

Forms that request information on a person’s race often request information on other characteristics including sex, age, and ethnicity as displayed in Table 5. These characteristics are useful in record matching as well as improving characteristics of existing records.

Table 5. Presence of Other Demographic Questions When a Race Question is Asked

Race Categories	Percentage of Time Present		
	Sex	Age	Ethnicity
<i>Disaggregated for Asian Americans and NHPI</i>	100.0%	100.0%	100.0%
<i>1997 OMB Categories and Undefined</i>	88.5%	83.9%	98.9%

V. Conclusion

In an effort to understand the coverage of Asian American and NHPI communities in federal administrative records, we reviewed data from OMB’s [Inventory of Approved Data Collections](https://www.reginfo.gov/public/do/PRAMain)⁴⁹ to determine the degree to which federal data

⁴⁹ <https://www.reginfo.gov/public/do/PRAMain>

collections not only cover Asian American and NHPI communities, *but also to what extent they provide disaggregated race data on these populations.*

Race data available from existing federal administrative data sources is far less prevalent than information such as name and address, especially when compared to survey and census data. There are approximately 342 million estimated form responses of the estimated seven billion (including collection forms representing an individual or an entire household useful for rostering) that contain race information. Approximately 208 million responses include disaggregated race for Asian Americans and NHPI and are solely sourced from survey data collections. When race data are collected on federal administrative data forms for an estimated 64 million responses, the categories of race are limited to the 1997 definitions required by OMB. Nevertheless, this information could be useful in multiple ways in survey and census operations. Federal data collections have promise for enhancing the decennial census and other Census Bureau survey operations. These data collections would be most useful to supplement the Census Bureau's existing universe of addresses and methods of contact, already created from a combination of administrative data sources and field operations, which could be employed in conducting censuses and surveys. Additionally, the inclusion of PII on federal data forms provides the ability to confirm or improve the attribution of an individual's physical location at a point in time.

Our subsequent research will explore the coverage and content of race information available from data collections conducted by state governments. In a final installment, we will develop recommendations on whether these data collections could be used to enhance statistics for the Asian American and NHPI communities, which specific data collections have the most promise, and how best they could be used.

Appendix A

Form Review Spreadsheet Column Definitions.

Variable Name	General Description	Valid types of information	Other Notes
Source Date	Date the record was captured	Date	
Source	Where we found the record	PRA-OMB (Paperwork Reduction Act Review)	
Government Level	Federal, State	Federal, State	
OMB Control No.	OMB Control number for PRA Review	OMB 4-digit Agency/Subagency Code: Request ID	
Agency/Sub	Name of Agency/Subagency	OMB Agency/Subagency Abbreviations	
Collection Agency if Different From Sponsor	The agency collecting the information if different from the submitting agency	Agency Abbreviations, defined by Demographic Analytics Advisors	
Title	Name of Data Collection	OMB Title	
Request Type	Type of review requested by the agency	OMB Types: New, Resubmission, Revision, Extension, Reinstatements	
Date Received	Date PRA Review Paperwork received by OMB	Dates defined by OMB	
Concluded Date	Date PRA Review Completed by OMB	Dates defined by OMB	
Conclusion Action	OMB's PRA Final Determination	Approved with or without changes, Not Subject to PRA, Withdrawn, Preapproved, Improperly submitted, Improperly submitted but continue	
Current Expiration Date	Date at which OMB terminates data collection approval unless re-requested	Dates defined by OMB	

No. of ICs	Number of Information Collection Activities defined by the agency	Number of collections submitted by agency	
No. of Forms	Number of Information Collection Forms defined by agency	Forms submitted by an agency for OMB Review	
OMB Status	Status of OMB form approval	Active, Under Review, Historically Active for Benefits, Mandatory, Voluntary	
Type of Data	Demographic Analytics Advisors: Type of data collection	Statistical, ADREC	
OMB Type Class	Type of data collection determined by OMB	Statistical, ADREC	
Estimated Number of Records	Estimated number of records provided to OMB for PRA review	Estimated number of records to be collected	
Why Collected	Demographic Analytics Advisors Determination or that of OMB	Notes	
Status	Demographic Analytics Advisors review status	Reviewed, Not Reviewed	Internal control note. All ICRs were analyzed. "Reviewed" denotes a second round of review for quality control.
Final Assessment	Demographic Analytics Advisors based on all reviews	In Scope, Out of Scope, Roster only	
Intermediate Assessment	Demographic Analytics Advisors categorization of ICRS	Categories 1 through 4	Category 1: Linkable with location, includes race data, and filter=eligible. Category 2: Linkable, includes race data, and filter=eligible. Category 3: Linkable, does not include race data, filter=eligible, and estimated number of responses are less than 250,000. Category 4: Out of Scope (not included in categories 1-3). Algorithm based.
Access Filter	Demographic Analytics Advisors determination that after review information collections should not be included	Eligible, National Security, Universe Issue	Removal of National Security information collections and ICRs with universe or form issues. Algorithm based.

Universe/Form Note	Demographic Analytics Advisors	Notes	Includes duplication of universe with other data collections (ex. Current Population Survey Supplements); if there was not enough information to determine the contents of the form; or several other issues that disqualified the information collection from this study
Initial Suitability Assessment	Demographic Analytics Advisors suitability assessment	In Scope, Out of Scope, Potentially Usable	In Scope if linkable or linkable with location and race is present, Potentially Usable if linkable, Otherwise Out of Scope. Algorithm based.
Linkable	Demographic Analytics Advisors linkage assessment	Linkable, Not Linkable, Linkable with Location	Linkable if Name or SSN is present, Linkable with location if Name or SSN is present and Address or Telephone Number is present, otherwise Out of Scope. Algorithm based.
Name	Demographic Analytics Advisors determination	Does name exist on the form?	
SSN	Demographic Analytics Advisors determination	Does SSN exist on the form?	
Address	Demographic Analytics Advisors determination	Does address information exist on the form?	
Telephone	Demographic Analytics Advisors determination	Does telephone information exist on the form?	
Email	Demographic Analytics Advisors determination	Does email information exist on the form?	
Age	Demographic Analytics Advisors determination	Does age information exist on the form?	Age or Date of Birth
Sex	Demographic Analytics Advisors determination	Does sex information exist on the form?	
Race	Demographic Analytics Advisors determination	Does race information exist on the form?	
Ethnicity	Demographic Analytics Advisors determination	Does ethnicity information exist on the form?	

Age Detail	Demographic Analytics Advisors determination	What age data is available?	
Race Detail (at finest level of disaggregation)	Demographic Analytics Advisors determination	What race detail exists on the form?	
Ethnicity Detail	Demographic Analytics Advisors determination	What ethnicity detail exists on the form?	
Sex Detail	Demographic Analytics Advisors determination	What sex/gender detail exists on the form?	

Appendix B. Related Readings Not Cited in This Report

Administrative Records Development

Towards a U.S. Population Database From Administrative Records

<https://www.census.gov/content/dam/Census/library/working-papers/1996/adrm/km9601.pdf>

Expansion Of Administrative Records Uses At The Census Bureau: A Long-Range Research Plan

https://www.researchgate.net/publication/237755223_EXPANSION_OF_ADMINISTRATIVE_RECORDS_USES_AT_THE_CENSUS_BUREAU_A_LONG-RANGE_RESEARCH_PLAN

Creating a Longitudinal Data Infrastructure at the Census Bureau

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/2015-alexander.pdf>

Name Census: United States Demographic Data

<https://namecensus.com/>

Linkage and Rosters

Digitizing Hand-Written Data with Automated Methods: A Pilot Project Using the 1990 U.S. Census

<https://www.census.gov/library/working-papers/2021/econ/digitizing-hw-data.html>

The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf>

Estimating Record Linkage False Match Rate for the Person Identification Validation System

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-02.pdf>

Evaluating the Master Address File—Auxiliary Reference file (MAF-ARF) as a Potential Respondent Retention Source

<https://www2.census.gov/ces/wp/2018/CARRA-18-12.pdf>

Playing with Matches: An Assessment of Accuracy in Linked Historical Data

<https://www.census.gov/content/dam/Census/library/working-papers/2016/adrm/carra-wp-2016-05.pdf>

Assessing Coverage and Quality of the 2007 Prototype Census Kid-Link Database

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-07.pdf>

Matching Addresses between Household Surveys and Commercial Data

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-04.pdf>

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

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-14.pdf>

Creating Linked Historical Data: An Assessment of the Census Bureau's Ability to Assign Protected Identification Keys to the 1960 Census

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-12.pdf>

The Decennial Census Digitization and Linkage Project

<https://www.census.gov/content/dam/Census/library/working-papers/2019/econ/dcdl-workingpaper.pdf>

Linking the 1940 U.S. Census with Modern Data

<https://www.tandfonline.com/doi/full/10.1080/01615440.2018.1507772>

Public Attitudes Toward the Use of Administrative Records in the U.S. Census: Does Question Frame Matter?

<https://www.census.gov/content/dam/Census/library/working-papers/2012/adrm/rsm2012-04.pdf>

Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIP)

<https://www.census.gov/content/dam/Census/library/working-papers/2011/adrm/ssm2011-03.pdf>

Using Administrative Record Data to Evaluate the Quality of Survey Estimates

<https://www.census.gov/content/dam/Census/library/working-papers/2010/adrm/rsm2010-10.pdf>

Proxy Reports: Results from a Record Check Study

<https://www.census.gov/content/dam/Census/library/working-papers/2010/adrm/rsm2010-09.pdf>

2018 American Community Survey Research And Evaluation Report Memorandum Series # ACS18-RER-07

https://www.census.gov/content/dam/Census/library/working-papers/2018/acs/2018_Clark_01.pdf

Full Report: Comparisons of Administrative Record Rosters to Census Self-Responses and NRFU Household Member Responses

<https://www.census.gov/library/working-papers/2023/dec/dec-wp-2023-01.html>

Evaluating Administrative Records as a Potential Sample Frame for the National Survey of College Graduates

<https://www2.census.gov/ces/wp/2018/CARRA-18-14.pdf>

Basic Demographic Studies

Measuring All-Cause Mortality with the Census Numident File

<https://www.census.gov/content/dam/Census/library/working-papers/2021/econ/wp-measuring-all-cause-mortality.pdf>

Measuring U.S. Fertility using Administrative Data from the Census Bureau

<https://www.census.gov/content/dam/Census/library/working-papers/2021/econ/chck2-adep-working-paper.pdf>

Coverage of Children in the American Community Survey Based on California Birth Records

<https://www2.census.gov/library/working-papers/2023/adrm/ces/CES-WP-23-46.pdf>

The Opportunities and Challenges of Linked IRS Administrative and Census Survey Records in the Study of Migration

<https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-06.pdf>

The Use of Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census

<https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-05.pdf>

Likely Transgender Individuals in U.S. Federal Administrative Records and the 2010 Census

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-03.pdf>

Race and Ethnicity Studies

Individual Changes in Identification with Hispanic Ethnic Origins: Evidence from Linked 2000 and 2010 Census Data

<https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-08.pdf>

Reporting of Indian Health Service Coverage in the American Community Survey

<https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-04.pdf>

Foreign-Born and Native-Born Migration in the U.S.: Evidence from IRS Administrative and Census Survey Records

<https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-07.pdf>

Using Linked Data to Investigate True Intergenerational Change: Three Generations Over Seven Decades

<https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-09.pdf>

When Race and Hispanic Origin Reporting are Discrepant Across Administrative Records and Third Party Sources: Exploring Methods to Assign Responses

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-08.pdf>

America's Churning Races: Race and Ethnic Response Changes between Census 2000 and the 2010 Census

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-09.pdf>

Dynamics of Race: Joining, Leaving, and Staying in the American Indian/Alaska Native Race Category between 2000 and 2010

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-10.pdf>

An Outside View: What Do Observers Say About Others' Races and Hispanic Origins?

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-05.pdf>

When Race and Hispanic Origin Reporting are Discrepant Across Administrative Records and Third Party Sources: Exploring Methods to Assign Responses

<https://www.census.gov/library/working-papers/2015/adrm/carra-wp-2015-08.html>

Assimilation and Coverage of the Foreign-Born Population in Administrative Records

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-02.pdf>

Exploring Administrative Records Use for Race and Hispanic Origin Item Non-Response

<https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-16.pdf>

Evaluating Race and Hispanic Origin Responses of Medicaid Participants Using Census Data

<https://www.census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-01.pdf>

Health-Related Administrative Records Studies

<https://www.census.gov/content/dam/Census/library/working-papers/2006/adrm/prevost-copafs.ppt>

Response Error and the Medicaid Undercount in the Current Population Survey

<https://www.census.gov/content/dam/Census/library/working-papers/2016/adrm/carra-wp-2016-11.pdf>

Medicare Coverage and Reporting: A Comparison of the Current Population Survey and Administrative Records

<https://www.census.gov/content/dam/Census/library/working-papers/2016/adrm/carra-wp-2016-12.pdf>

Response Error and the Medicaid Undercount in the Current Population Survey

<https://www.census.gov/content/dam/Census/library/working-papers/2016/adrm/carra-wp-2016-11.pdf>

Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, the National Center for Health Statistics, the Administration for Healthcare Research and Quality, and the U.S. Census Bureau

<https://www.census.gov/content/dam/Census/library/working-papers/2007/adrm/snacc-phase-i-full-report.pdf>

<https://www.census.gov/content/dam/Census/library/working-papers/2008/adrm/snacc-phase-ii-full-report.pdf>

<https://www.census.gov/content/dam/Census/library/working-papers/2008/adrm/snacc-phase-iii-full-report.pdf>

<https://www.census.gov/content/dam/Census/library/working-papers/2009/adrm/snacc-phase-iv-full-report.pdf>

<https://www.census.gov/content/dam/Census/library/working-papers/2010/adrm/snacc-phase-v-full-report.pdf>

Misreporting Health Insurance Status: Medicaid Enrollees in the Medical Expenditure Panel Survey, 20031

<https://www.census.gov/content/dam/Census/library/working-papers/2009/adrm/lynch-aapor-08-05-2009.pdf>

Is There an Undercount of Medicaid Participants in the ACS Content Test?

<https://www.census.gov/content/dam/Census/library/working-papers/2009/adrm/medicaid-participants-acs-content-test.pdf>

Medicaid Underreporting in the CPS: Results from a Record Check Study

<https://www.census.gov/content/dam/Census/library/working-papers/2007/adrm/rsm2007-40.pdf>

Preliminary Analysis of Medicaid Enrollment Status in the Current Population Survey

<https://www.census.gov/content/dam/Census/library/working-papers/2007/adrm/preliminary-analysis-mah-2007.pdf>

Income-Related Studies

SNAP Receipt in SIPP: Using Administrative Records to Evaluate Data Quality

<https://www.census.gov/library/working-papers/2022/demo/SEHSD-wp2022-22.html>

The Summary Earnings Record and Detailed Earnings Record Extracts

<https://www.census.gov/library/working-papers/2021/econ/earnings-record-extracts.html>

When and Why Does Nonresponse Occur? Comparing the Determinants of Initial Unit Nonresponse and Panel Attrition

<https://www2.census.gov/library/working-papers/2023/adrm/ces/CES-WP-23-44.pdf>

Evaluating Administrative Records to Inform Measurement Error Properties of National Survey of College Graduates Estimates: An Analysis of the NSCG-LEHD Earnings Ratio

<https://www2.census.gov/ces/wp/2018/CARRA-18-13.pdf>

Evaluating Administrative Records to Inform Measurement Error Properties of National Survey of College Graduates Estimates: Employment History and Firm Characteristics

<https://www2.census.gov/ces/wp/2018/CARRA-18-11.pdf>

Is Subsidized Childcare Associated with Lower Risk of Grade Retention for Low-Income Children? Evidence from Child Care and Development Fund Administrative Records Linked to the American Community Survey

<https://www.census.gov/content/dam/Census/library/working-papers/2017/adrm/carra-wp-2017-06.pdf>

Capturing more than poverty: School free and reduced-price lunch data and household income

<https://www.census.gov/content/dam/Census/library/working-papers/2017/adrm/carra-wp-2017-09.pdf>

Tax Preparers, Refund Anticipation Products, and EITC Noncompliance

<https://www.census.gov/content/dam/Census/library/working-papers/2017/adrm/carra-wp-2017-10.pdf>

The Use and Misuse of Income Data and Extreme Poverty in the United States

<https://www.census.gov/content/dam/Census/library/working-papers/2019/econ/adeq-wp-2019-02.pdf>