

Assessing Availability and Quality of Administrative Records for Asian Americans and Native Hawaiians and Pacific Islanders: Recommendations for the Future



I. Introduction

This is the final report of our three-part series focused on the potential role administrative records can play in improving enumeration for both the Asian American and the Native Hawaiian and Pacific Islander (NHPI) populations in decennial censuses as well as in other Census Bureau products. The focus of these reports is on the coverage of data on race—and specifically disaggregated race—in administrative records, and the potential limitations and benefits of using these records within an enumeration of these populations. The first report, [Assessing Availability and Quality of Administrative Records for Asian Americans and Native Hawaiians and Pacific Islanders: Introduction and Federal Database Analysis](#), focuses on the coverage of the Asian American and NHPI populations in federal administrative records and the attributes that can be assigned to these records. The second report, [Assessing Availability and Quality of Administrative Records for Asian Americans and Native Hawaiians and Pacific Islanders: State Database Analysis](#), focuses on the same topics, but for state-level administrative records.

Each of the above reports provide some reasons for optimism as well as significant cautions about the use of administrative records in decennial census operations. This final report focuses on unpacking potential upsides and downsides, while outlining recommendations for both the Census Bureau and their myriad stakeholders as we approach the 2030 Census. This report begins with a short recap of our previous findings as well as conclusions from other research about the coverage of race in administrative records, with a specific focus on the Asian American and NHPI communities. We then briefly cover the changes in Statistical Policy Directive 15 (SPD 15) that impact the collection of race and ethnicity data both in the decennial census and in administrative records. Finally, we share our cautions and recommendations about the usage of administrative records in census operations, with a focus on what we believe the Census Bureau and stakeholders should prioritize in the lead up to the 2030 Census and beyond.

II. Findings: Federal and State Administrative Records Coverage on Race¹

The Census Bureau has been researching and implementing the use of administrative records, at least for some census operations, for the last several decades. These uses include Master Address File creation and quality checks, quality checks of the enumeration, and some data imputation. Overall, Census Bureau research and development has shown that these records can be used to reach similar measures of population totals as the decennial census, but major limitations remain in their use for the measurement of race and ethnicity.

Indeed, research over the past decade at the Census Bureau has shown the potential benefits and limitations of using race and ethnicity data from administrative records and third-party data, especially when race or Hispanic origin data might be missing on a census or survey (i.e., the approximately 3%² of

¹ While we are focused here on race, many of the same concerns and benefits would hold for ethnicity as well.

² Three percent represents over 10 million people when enumerating the full US population over the time period of the research.

the enumerated population that has race data that are missing or disagree across records as well as the approximately 3% of the population that has missing data or data disagreements across records for Hispanic origin).³ This research, focusing on what the Census Bureau calls the “Best Race” file, outlines the business rules for how the Census Bureau applies race and Hispanic Origin to records using administrative and third-party data.⁴ Overall, even when decennial census and administrative data⁵ sources can be matched on race and Hispanic Origin characteristics, disagreements between the two datasets are higher for minority groups. Individuals aged 0-17 also have higher levels of disagreement. Finally, the Census Bureau’s research only focuses on aggregated race categories, therefore not answering the question of the potential applicability and quality of these data for disaggregated race categories.

Other research during and after the 2020 Census shows similar findings. For instance, the 2020 Census real-time administrative record simulation reported a population that was over 2% higher than the decennial census for the total population. However, it did not try to take into account disaggregated race, and showed lower counts for several race groups, including the non-Hispanic Asian population.⁶ Finally, this research, unlike the Best Race research above, used data on citizenship status from administrative records. This should not be a part of the Census Bureau collection to ensure privacy, to build trust in all communities, and because those data points are often not current in the administrative records.

Our own research has also confirmed these limitations, while adding the nuance that the lack of data disaggregation for racial subgroups creates even greater limitations for administrative data use. In our work on federal administrative data sources, we showed that the vast majority of data are collected in the 1997 Office of Management and Budget (OMB) minimum categories, without disaggregated racial subgroup options. Despite this lack of data disaggregation, these data can still be useful in a variety of ways. This is particularly important as even though changes based on the 2024 revision of SPD 15 are intended to increase the collection of disaggregated data, as we explain below, this is unlikely to occur quickly enough for use in the 2030 Census.

State-level administrative records are more of a mixed bag. As we showed in our analysis in the second report in this series, some states do collect disaggregated race data on some administrative records

³ Noon, James. “Revisiting Methods to Assign Responses When Race and Hispanic Origin Reporting Are Discrepant across Administrative Records and Third Party Sources.” *Census.Gov*, U.S. Census Bureau, May 2024, www.census.gov/library/working-papers/2024/adrm/CES-WP-24-26.html.

⁴ Ennis, Sharon R., et al. “When race and Hispanic origin reporting are discrepant across Administrative Records and third party sources: Exploring methods to assign responses.” *Statistical Journal of the IAOS*, vol. 34, no. 2, 17 May 2018, pp. 1979–1989, <https://doi.org/10.31235/osf.io/5xhh6>.

⁵ The Census Bureau has more recently moved from the use of the term “administrative records” to “administrative data.” Generally these terms are used to discuss the same sources.

⁶ Brown, J. David, et al. “Real-Time 2020 Administration Record Census Simulation.” *Census.Gov*, U.S. Census Bureau, 5 May 2023, www2.census.gov/programs-surveys/decennial/2020/program-management/evaluate-docs/EAE-2020-admin-records-experiment.pdf.

forms, while others do not. Further, some forms, such as birth and death records, are more useful in this area than others. Overall, the Census Bureau would benefit from using many of these data sources, some of which they already are. However, testing should be conducted to show the level of agreement between census and administrative record responses for disaggregated race categories, as the Census Bureau has already done for the current minimum reporting categories.

III. Impact of Changes to SPD 15

The March 2024 updates to SPD 15 have introduced significant changes in how data on race and ethnicity are collected. Two of the biggest changes are the introduction of a Middle Eastern and North African reporting group, and the combining of previous race and ethnicity questions into one combined question. Another key change is the collection of detailed—or disaggregated—race and ethnicity data by default. This shift acknowledges that aggregated data can mask important disparities within subgroups, thus hindering effective policymaking and fair resource allocation.

According to OMB: “The revised SPD 15 requires the collection of detailed data on race and ethnicity beyond the minimum categories, unless an agency determines that the potential benefit of the detailed data would not justify the additional burden to the agency and the public or the additional risk to privacy or confidentiality, and therefore requests an exemption from the Office of Management and Budget’s (OMB) Office of Information and Regulatory Affairs (OIRA). In those cases, Federal agencies must at least use the minimum categories and justify this determination in the agency’s PRA information collection review package.”⁷ This means that, unless an agency requests and is approved to collect aggregate data only, all administrative and survey data collections by federal agencies will include disaggregated race and ethnicity data.

This is a promising start that may prompt additional improvements. For example, states often comply with the same guidelines as federal agencies. Some of these are required changes, such as birth and death records that are required to meet the same federal minimum reporting standards (although states are given a longer period of time to comply). Other collections are left entirely up to the states, and many states make these decisions based on their data collection needs. However, states can often find it easier to point to federal rules in this process. Some states will even focus on specific populations with large communities in their state and decrease reporting for groups that they know have more limited presence.

However, even if all populations were sufficiently covered in administrative data sources, when it comes to implementation of the new SPD 15, it is unlikely that disaggregated race and ethnicity data will be available in time for testing and use in the 2030 Census, even assuming full implementation by the required timeline. These changes take time to put in place, as the following guidance given by OMB makes clear:

⁷ <https://spd15revision.gov/content/spd15revision/en/2024-spd15/question-format.html>

“The provisions of these standards are effective March 28, 2024 for all new record keeping or reporting requirements that include race and ethnicity data. All existing record keeping or reporting requirements should be made consistent with these standards through a non-substantive change request as soon as possible, or at the time they are submitted for extension or revision to the Office of Management and Budget’s (OMB) Office of Information and Regulatory Affairs (OIRA) under the Paperwork Reduction Act (PRA), but not later than March 28, 2029.”⁸

Indeed, there is a possibility that this 2029 deadline will not be met. While federal agencies conformed to the 1997 OMB race standards in their own data collections in a timely fashion, it took many years for those standards to be implemented in federal programs that collected information through state agencies.⁹ Because the standards did not directly apply to states, the reporting of race and ethnicity information was not always consistent. It is likely that the new 2024 SPD 15 race and ethnicity reporting standards will also take many years to be consistently implemented in state administrative records (for example, health and human services and education data). Despite the requirements set out in SPD 15, it is conceivable that some agencies will not be in full compliance by the stated deadline of March 28, 2029. Additionally, even if all federal agencies were to be in compliance by the deadline set by OMB, it would be too late for use in the 2030 Census. There would not be adequate time for the significant, necessary testing of how these changes in administrative records may impact the quality of the data. Additionally, multiple Census Bureau research projects have shown that the reporting of race and ethnicity in administrative records, even for the more limited 1997 OMB minimum reporting categories, can disagree with self-reporting in the decennial census. Understanding how these differences are playing out is key to confidently implementing the use of these administrative data in decennial census operations.

State and local government records, including birth and death records, will likely take even longer to fully implement, including properly disaggregating data. In fact, the last time SPD 15 was changed in 1997, it took over 20 years for all states to be in compliance with their collections for births and deaths.¹⁰ Therefore, while there may be greater availability of disaggregated race and ethnicity data from administrative records in the future, it is highly unlikely that it will be ready early enough to test its quality and usefulness in the 2030 Census. Even when new data collection policies are fully implemented, their impact on the collected data will take additional time to appear in subsequent datasets. For example, new forms for benefit applications that implement disaggregated race and ethnicity categories will likely only apply to new and/or renewing beneficiaries. This issue leads to the need for bridging data between the two sets of categories, which, if done correctly, allows for longitudinal analysis.

⁸ <https://spd15revision.gov/content/spd15revision/en/2024-spd15/effective-date.html>

⁹ <https://www.macpac.gov/wp-content/uploads/2023/03/Chapter-1-Medicaid-Race-and-Ethnicity-Data-Collection-and-Reporting-Recommendations.pdf>

¹⁰ https://www.cdc.gov/nchs/nvss/bridged_race/Documentation-BridgedPostcenV2020.pdf

IV. The Use of Administrative Records in Census Operations: Cautions and Recommendations

Given the above information, we have several cautions as well as recommendations for the use of administrative records in decennial census operations. As we noted in the first report:

“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.”¹¹

We begin first with our cautions, moving next to the areas we believe administrative records can be most helpful, and finishing with further research and policy recommendations to make the use of administrative records more effective in improving coverage for populations like Asian Americans and NHPs.

Cautions

First and foremost, it is important to state that the Census Bureau should not move to an administrative record-first or administrative record-only enumeration. This would be an extreme shift and is unlikely to happen, but it still bears mentioning. Relatedly, we heavily caution against the Census Bureau moving to replace or shorten Nonresponse Followup (NRFU) with the use of administrative records. Research has shown that household self-response provides higher quality data, especially on demographic characteristics,¹² and these higher quality data should not be lost, especially for the harder-to-enumerate populations often reached by NRFU.

Racial and ethnic disparities in administrative records have been documented in different contexts, such as in healthcare.¹³ Any disparities in administrative records would be carried over, and potentially exacerbated, if used especially excessively in the decennial census. This is particularly problematic for

¹¹ <https://www.advancingjustice-aajc.org/sites/default/files/2024-03/Administrative%20Records%20Report.pdf>

¹² See, for example, Salvo, Joseph J., et al. “Census 2020 why increasing *self*-response is key to a good count.” *Significance*, vol. 17, no. 1, 29 Jan. 2020, pp. 30–33. and https://www.census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_247.pdf

¹³ <https://www.advancingjustice-aajc.org/publication/advancing-justice-aajc-submits-comments-response-announcement-pacaahmpi-meeting-and>

minority population groups who are simultaneously more likely to be missed in the census and less likely to be captured in administrative records.¹⁴

Additionally, due to the lack of standardization across administrative databases on how race and ethnicity data are reported and collected, a large portion of our community could potentially be missed by relying solely on administrative data. In many datasets, “Asian” responses as well as “NHPI” responses are lumped into an “Other” category, making it impossible to determine which individuals would identify as Asian American or NHPI. Furthermore, outside of the data produced by the Census Bureau, detailed data on Asian American and NHPI subgroups are generally unavailable.¹⁵ Finally, there are some administrative records that are not collected through self-response but rather through observation or proxy response and are thus of lower quality for characteristics such as race and ethnicity. Thus, reliance on administrative records may, at best, provide occasional data on Asian Americans and NHPs and rarely, if ever, on detailed subgroups.

Any administrative records used by the Census Bureau should also be clear on the time period in which the data were collected. The decennial census is unique in that it measures the location and characteristics of the entire resident population as of a statutorily established date: April 1st of the census year. State administrative records generally have no such benchmark date. Therefore, information gathered from program participants or benefit recipients could represent wide and inconsistent windows of time that diminish the usefulness of the information to supplement decennial census statistical products. This must be taken into consideration when assessing whether to acquire a particular data source and/or how to utilize it, with administrative records being used for enumeration purposes only if that time period is relevant to the statutory April 1, 2030 date of enumeration. This is especially important for communities that are highly mobile.

Of particular concern is the potential use of commercial sources for data, as commercial data have been found to be of lower quality than governmental data with respect to the Asian American and NHPI communities.¹⁶ Specifically, commercial data have been found to potentially be useful for certain characteristics, such as age and sex, but are of much more dubious quality for race and ethnicity—that is, race and ethnicity characteristics from commercial sources match decennial census race and ethnicity data at a lower rate than other characteristics (like age and sex).¹⁷ This is due to several potential factors. A primary concern is that commercial data on race and ethnicity are often modeled, not based on self-response. Moreover, most do not even model a multiracial population, which skews their modeled race data, particularly for groups (such as NHPs) with higher numbers of multiracial individuals.

Finally, any use of administrative records should be transparent about privacy, and potential harm to respondents needs to be proactively taken into account. The Census Bureau should be clear about any

¹⁴ Brown et al, “Real-Time 2020 Administrative Record Census Simulation.”

¹⁵ <https://www.advancingjustice-aajc.org/sites/default/files/2024-03/Administrative%20Records%20Report.pdf>

¹⁶ https://www.census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_247.pdf

¹⁷ Ibid

and all administrative records that are used in any way for decennial census operations, how they are used, how privacy of those data are protected, and how self-responses from the decennial census may be merged with administrative data sources for further use.

Recommendations for Usage

Given these cautions, how could administrative data best be used in the decennial census? There are some areas where administrative records should be used, others where limited application may be useful, others where administrative records should be a last resort, and others where they should not be used at all.

In terms of instances where administrative records should be used, the use of administrative records continues to be absolutely essential in developing the Master Address File (MAF) to improve population coverage and support self-response in the decennial census. The MAF is the backbone of Census Bureau demographic surveys and censuses, as it provides the universe of addresses and physical locations for the delivery of questionnaires and the conduct of field operations, as well as providing geographic attributes required for data tabulation. Any additional administrative records that can improve the MAF, while still protecting respondent privacy, should be used to ensure everyone is counted. Simply put, if an address is not in the MAF, that household will not be counted.

Another way that administrative data could be usefully applied in a limited manner is to improve the chances that Asian Americans and NHPI populations are contacted during the decennial census, as well as informing the Census Bureau of the mode and language that are most likely to get a response from these groups. This idea is closely tied to the use of administrative and other third-party data sources for Get Out the Count (GOTC) operations, both by the Census Bureau as well as by outside groups. Again, as long as these data are sufficiently privacy protected, and the methods of outreach are tested and measured for success, this could be a beneficial use of administrative data to increase self-response rates and therefore the overall quality of decennial census data.

One “last resort” use of administrative records is to fill in missing data due to nonresponse, including adding data with either detailed or aggregate race information. To be clear, self-identification of race and ethnicity is of vital importance to our communities, therefore this data should only be used after all attempts at NRFU have failed. This conclusion is based on the quality differentials found between self-response and other modalities.¹⁸ Because of this, prior decennial census and American Community Survey data may be a good initial option to fill in missing data, where available. Further, if the Census Bureau uses administrative records to fill in information in this way, it is our recommendation that they use something like the best race algorithm that they have already studied, making improvements to the

¹⁸ Specifically, the number of NRFU attempts should not fall below the standard of six contacts used in the 2020 Census without further research that shows such a reduction would not harm self-response rates in historically or persistently undercounted communities. Indeed, further research should be undertaken to build a robust NRFU outreach plan that may include more contacts and/or additional effective ways to engage the population that ensures more households are able to self-respond during NRFU.

algorithm only with substantial input from stakeholders. This should include extensive study of which records are most likely to have high-quality race and ethnicity reporting (including detailed race reporting), the potential impact of using probability models, and qualitative input from the groups most likely to be impacted by inclusion of these data sources.

Although administrative data could prove useful in the ways described above, we remain adamant that administrative records *should not* be used to conduct an administrative records-only census. The decennial census is paramount in ensuring equal political representation and voting rights. Individuals must be provided the ability to self-respond through the decennial census, and, as we noted above, these self-responses are generally found to be of higher quality than the use of administrative records.

In order to best operationalize these recommendations, each administrative record should be evaluated for how it may help improve census data coverage, and this information should be made public. For instance, there may be administrative records that have no race or ethnicity data, but which would help to address issues in the MAF. These data could still be extremely important to the Asian American and NHPI communities by improving the overall coverage of the community through improvements to the underlying address frame. This process will allow prioritization as well as transparency in the use of data; a record that could have a moderate impact on each of the modes of utility could be prioritized, together with a record that has a significant impact on one of the modes (for instance by substantially improving the MAF).

Additional Considerations and Recommendations

In addition to the above recommendations on how the Census Bureau should prioritize the inclusion of and research into administrative records, we also believe the Census Bureau and other stakeholders should consider the following when determining whether and how to utilize administrative records as part of data collection and reporting:

1. Developing a quality scorecard for each administrative record source
2. Accounting for the uneven population and geographic coverage of administrative records
3. Ensuring transparency in data collection and sharing

Below, we provide additional context on each of these points.

Developing of a quality scorecard for each administrative records source

The use of administrative records for statistical applications requires an understanding of the goals of the organization or agency producing the original source records in order to better assess the strengths and weaknesses of the data. We therefore recommend the bureau develop a quality scorecard for each source of administrative records they have acquired. At the very least, this would include an overall assessment of each data source as to its fitness for use for various purposes, such as use for the MAF, improving item or household non-response in surveys and censuses, or use in the Population Estimates

Program.¹⁹ This scorecard should also contain metadata including (but not limited to) the methods used to collect, purge, and edit individual records as well as the definition of each variable and its importance to the collecting agency’s mission. Appendix A includes an example of what a scorecard like this might look like. Such a scorecard should cover eight areas at a minimum:

1. General source information
2. Timeliness of the data
3. Completeness of the data
4. Validity of the data for specific uses
5. Consistency of the data
6. Integrity of the data
7. Accuracy of the data
8. Priority use cases

Prior to use, these scorecards should be publicly available for all data sources the Census Bureau plans to use in its specific operations, including the decennial census.²⁰ As a starting point, we are happy to share with the Census Bureau the database of federal and state administrative record sources that we have already reviewed.

Accounting for the uneven population and geographic coverage of administrative records

As we saw in our report on state administrative records, not every state collects the same administrative data, which can be a concern depending on the purpose for which an administrative record is being used. We are also concerned that not all states collect data of comparable quality and accuracy, even when they are collecting the same information. Different program implementation protocols and eligibility requirements for program participation further compromise the consistency of data from state to state. Even if every state follows the same standards on the collection and categorization of data for a particular topic, the method of collection may impact the quality and availability of that information.

For instance, the “Application for Food and Cash Assistance” for the state of Washington asks applicants to write in their race (see below):

¹⁹ Item nonresponse occurs when a respondent completes part of the questionnaire but does not answer one or more individual questions. <https://www2.census.gov/programs-surveys/sipp/guidance/spd/types-missing.pdf>

²⁰ The grading process for whether and how to use this data in NRFU for the decennial census should be focused in part on the availability of detailed race data.

Figure [x]. Excerpt of Application for Food and Cash Assistance, Washington State²¹

18. List everyone in your household even if you are not applying for them (attach additional sheets, if needed).

NAME (FIRST, MIDDLE, LAST)	GENDER	HOW IS THIS PERSON RELATED TO YOU?	DATE OF BIRTH	CHECK IF YOU WANT BENEFITS FOR THIS PERSON	OPTIONAL FOR NON-APPLICANTS			
					SOCIAL SECURITY NUMBER	CHECK IF U.S. CITIZEN	RACE (SEE SAMPLES BELOW)	TRIBE NAME (For American Indians, Alaska Natives)
		Myself		<input type="checkbox"/>		<input type="checkbox"/>		
				<input type="checkbox"/>		<input type="checkbox"/>		
				<input type="checkbox"/>		<input type="checkbox"/>		
				<input type="checkbox"/>		<input type="checkbox"/>		
				<input type="checkbox"/>		<input type="checkbox"/>		
				<input type="checkbox"/>		<input type="checkbox"/>		
				<input type="checkbox"/>		<input type="checkbox"/>		

19. My ethnic background is Hispanic or Latino: Yes No

Race and Ethnic background information is voluntary and will not affect eligibility or benefit amounts. This information is used to assure program benefits are distributed without regard to race, color, or national origin. For Food Assistance the USDA requires us to answer for you if no information is provided. We will select "unreported" if you don't want to answer. **Race examples:** White, Black or African American, Asian, Native Hawaiian, Pacific Islander, American Indian, Alaska Native, or any combination of races.

In comparison, a similar application for the state of Utah gives disaggregated race options for select Asian and NHPI groups:

Figure [x]. Excerpt of Application for SNAP, Financial Assistance, Child Care, and Medical Assistance²²

HOUSEHOLD AND GENERAL INFORMATION										
4. List everyone who is living in your household and applying for benefits:										
First and Last Name	Social Security # ¹	Birth Date	U.S. Citizen/ Eligible Non-Citizen Yes/No	Gender M / F	Relationship	Utah Resident Yes/No	Utah Resident Since ² (ex: 07/14/13)	Race ^{3,6}	Ethnicity ^{4,6}	Marital Status ⁵
					Self					

¹ Social Security number and Citizenship information are only needed for the people applying for benefits. If someone wants help getting a Social Security number, call 800-772-1213 or visit socialsecurity.gov. TTY users should call 800-325-0778. A Social Security number is not required for Child Care. Eligibility for Child Care will not be denied due to not providing a Social Security Number.

² Utah Resident is optional for all programs

³ Race (optional): AI = American Indian or Alaska Native (For medical applicants only, complete Attachment A)
 GC = Guamanian or Chamorro ASI = Asian Indian CH = Chinese JA = Japanese KO = Korean
 OPI = Other Pacific Islander FI = Filipino VI = Vietnamese AS = Asian OA = Other Asian
 BL = Black or African American SA = Samoan NH = Native Hawaiian OT = Other WH = White

⁴ Ethnicity (optional): N = Not Hispanic, Latino or Spanish Origin M = Mexican MA = Mexican American CH = Chicano/a
 PR = Puerto Rican CU = Cuban AH = Another Hispanic, Latino or Spanish Origin OT = Other

⁵ Marital Status is not required for SNAP

⁶ For SNAP and Medicaid: You do not have to give us racial/ethnic information. If you do not want to give us this information, it will have no effect on your case.

These examples also show that administrative records often target specific populations; in this case, individuals with lower income in need of specific services. Indeed, many administrative data sources

²¹ <https://www.dshs.wa.gov/sites/default/files/forms/pdf/14-001.pdf>

²² <https://jobs.utah.gov/forms/61app.pdf>

only cover subsets of the population. Even if these sources intend to collect information for the majority of the population, it does not mean that all relevant individuals actually participate in the program. For instance, although most children would be covered in public school enrollment data, the increase in home, private, and charter schooling may lead to differences in reporting between states.

We therefore recommend that the Census Bureau takes into consideration whether a particular type of administrative data source is available for all 50 states and equivalent areas, and ensure they maintain comparable quality across the United States. At the same time, there may be certain communities where certain data sources from targeted states may work best to accurately include and reflect specific communities; these cases should be assessed individually to ensure equity and accuracy. For example, the Bureau should consider using Supplemental Nutrition Assistance Program (SNAP), Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), and Temporary Assistance for Needy Families (TANF) data or other data sources that provide coverage on communities that are likely missing from Internal Revenue Service (IRS) records from states where the data quality is good, even if the same level of quality is not available across all 50 states. Additionally, the Bureau should place a high priority on acquiring vital records such as individual birth certificate information to address overall coverage as well as the persistent undercount of young children.

Indeed, administrative data sources may be helpful in enhancing the coverage of certain populations. For example, people experiencing homelessness are highly concentrated in a handful of states; the Bureau should prioritize getting quality and consistent data on this community from those specific states rather than applying a rigid rule of nationwide availability, consistency, and quality. Working with local and state governments to gain access to these data in a safe and secure way may help to improve the counts for these populations. Another example of this sort of very specific data is the Alaska Permanent Fund Dividend file. These data could be useful in assisting with a fair and accurate count of Alaskan Natives, but would only be relevant for one of the 50 states. This approach is similar to the Census Bureau's use of different enumeration modes across the country in decennial census operations. In any event, when using these administrative records, the Census Bureau should use the most detailed information available, even if those details are not available for all jurisdictions. This is especially important when records can be used to help improve missing data. The Census Bureau should then be clear about how these cases are coded in more detailed race tables.

In sum, we recommend that the Census Bureau take uneven data distribution into consideration when assessing how to use any administrative record to ensure that its use will not exacerbate historical and persistent inequities in count accuracy for the decennial census and other Census Bureau surveys. The Census Bureau should especially exercise great caution before using these governmental records in statistical products that have legal, policy, and funding purposes as any exacerbation of historical and persistent statistical disparities could result in direct harm to already marginalized communities. Furthermore, data that respondents might consider sensitive—such as their citizenship status, income, and household size (as well as race and ethnicity)—might be less reliable depending on the individual's own circumstances and the broader political environment, both nationally and in each state. This may yield administrative records of uneven and even dubious quality.

Ensuring transparency in data collection and sharing

In order to gain the trust and confidence of interested stakeholders and respondents, the Census Bureau must be very clear about which administrative data sources and data points are being collected and how they are to be used. This transparency should include how well administrative records match to survey responses. The race and ethnicity information provided through administrative data or through self-response represents the first step in developing datasets with accurate race and ethnicity characteristics. The next step involves how records are edited, and the Bureau needs to provide greater transparency on how respondents are attributed a race and ethnicity or detailed race category.²³ For example, if in the 2020 Census a person checked the box that they considered their race Asian American or NHPI but also wrote in under “Other Asian” or “Other NHPI” that they were from England, their response would have been edited to multi-racial (Asian or NHPI and White). However, this process conflates nationality with race, and makes assumptions about the respondent’s self-identification. What happens when an administrative record considers that person as Asian American alone or NHPI alone? Are administrative records considered a non-match, or could they be used in the editing process as a confirmation that this person was intending to self-identify as a single race?

Additionally, it is important that these race and ethnicity characteristics only be used for statistical purposes in accordance with Title 13. Increased and continued transparency on what data sources are brought into the Census Bureau, as well as any sharing of data back to those outside of the Census Bureau or further use of the administrative records beyond the initial purpose set out in their Memorandum of Agreement within the Census Bureau, must be clear and transparent. We recommend that the Census Bureau create a centralized website for detailing the data sources being utilized by the Bureau, in what manner they are to be used, and what data are to be shared back to those outside the Bureau. Additionally, this site should include the Memorandums of Agreement with the different partners from which they are obtaining administrative records. Finally, we recommend that the Bureau should not use administrative records beyond the purpose laid out in its Memorandum of Agreement. However, if there is any further use beyond the initial agreed-upon purpose, that information should also be shared on the centralized website.

V. Conclusion

As we have discussed, while there are opportunities to more efficiently and effectively use administrative data in the 2030 Census and beyond, overreliance on administrative records could harm the quality of the count. Further, this could lead to the deepening of existing disparities, particularly for historically undercounted communities, and compromise the accuracy of key demographic data such as race and ethnicity. Variability in the coverage, quality, and timing of administrative records, especially in areas such as detailed race and ethnicity data, requires a nuanced approach that carefully considers

²³ The Census Bureau should research the impact of their coding and tabulation rules using cognitive testing to ensure that when they are coding specific write-ins in specific locations on the form, they are doing so in a way that matches the respondent’s intent.

how, when, and where these records are used. Further, it is essential that administrative data not replace or shorten critical census operations like NRFU.

The thoughtful use of administrative records can play an important role in improving the census process, particularly in areas such as developing the Master Address File (MAF) and enhancing outreach efforts for underrepresented groups. However, administrative data should be used cautiously, and only as a supplement to traditional methods of self-response. To ensure equity and transparency, the Census Bureau must adopt a rigorous framework for evaluating and applying administrative records, including the development of quality scorecards and the transparent and clear communication of data usage policies. By striking the right balance between using innovative data sources and established census methods while also prioritizing historically marginalized groups, the Census Bureau can work toward a more accurate and equitable enumeration that reflects the diversity of the U.S. population.

Ultimately, the successful integration of administrative records into the decennial census will require ongoing research, stakeholder input, and careful monitoring in order to mitigate potential harms. Transparency in data collection, privacy protection, and the use of administrative records only for statistical purposes as laid out in Title 13 are essential to building trust and ensuring that the census accurately reflects all communities.

Appendix A

Scorecard example

Quality Domain	Response	Explanation (when necessary)
1.0 Source Information		
1.1 Data product name		
1.2 Who produces the data?		
1.3 Why does the producer collect the data?		
1.4 Who distributes the data?		
1.5 Is an agreement required to access and use the data?		
1.6 Are there policy constraints to using the data?		
1.7 Is there a cost to acquire the data?		
1.8 Does the data provider have reporting requirements for access, use, and disposition?		
2.0 Timeliness		
2.1 What is the reference period of the data source?		
2.2 When will the data source be available?		
2.3 Are data available when needed?		
2.4 Are the data produced on a recurring basis?		
2.5 Are the data produced on a consistent schedule?		
3.0 Completeness		
3.1 What is the geographic coverage of the data source?		
3.2 What is the population coverage of the data?		
3.3 How many records are typically included in the file?		
3.4 What information domain(s) does the file contain?		
4.0 Validity		
4.1 For what organizational purpose could the data be used?		
4.2 Is there documentation about the data collection process?		
4.3 Are certain variables more important to the data producer?		
4.4 Does the producer edit or suppress data fields on specific records and is there documentation?		

4.5 What variables need to be transformed for the recipient's organizational purpose?

4.6 How are variable values transformed and are they documented?

5.0 Consistency

5.1 Is geographic coverage consistent over time?

5.2 Is the population coverage consistent over time?

5.3 Are data collection processes consistent over time?

5.4 Have specifications for data variables changed over time?

5.5 Do duplicate records exist on the file?

5.6 Does the provider purge data on a regular basis?

6.0 Integrity

6.1 Are there security requirements involved in data transfer, intake, use, and disposition?

6.2 Is there documentation for the supplied file?

6.3 Were the number of records received consistent with expectations?

6.4 Is the format and contents of the file consistent with expectations?

6.5 Are files backed up once received?

6.6 Are files backed up once processed?

6.7 Can users change data without documentation and approval?

6.8 Does a change management process exist for processing and documentation alterations?

7.0 Accuracy

7.1 What is the rate/number of missing observations by variable?

7.2 Are variable values within expected range(s)?

7.3 Are there known issues with data quality?

7.4 Can identifying information be used to link this data source to others?

7.5 What is the rate of success in linking to other data sources?

7.6 Are there specific characteristics for records that do not link to other data sources?

8.0 Priority Use Cases

8.1 Does the source provide coverage in key housing types and geographies for MAF Improvement (e.g., accessory dwelling units, non-postal addresses, tribal lands, etc.)?

8.2 Is there information that will help improve the chances of enumeration (e.g., preferred modes of response, languages spoken, etc.)?

8.3 Is there detailed race and/or ethnicity data included?