



Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities

Key Results

- Medicare's enrollment data on race and ethnicity have inaccuracies and other limitations, particularly for those identified as American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic.
- The inaccuracies stem from deficiencies in the source data. The use of an algorithm also has drawbacks.
- Inaccuracies in Medicare's enrollment data limit the ability to assess and address health disparities.
- Medicare should develop its own source of race and ethnicity data.

Why OIG Did This Review

The disparate impacts of the COVID-19 pandemic on various racial and ethnic groups have brought health disparities to the forefront. Health disparities are differences in health that adversely affect certain groups. People of color have been found to experience disparities in areas such as access to health care and quality of health care.¹ Such disparities have profound implications for the health and well-being of these individuals.

Medicare is an essential part of the Nation's health care system, with 66 million beneficiaries enrolled. The Centers for Medicare & Medicaid Services (CMS) has made advancing health equity a top priority. Ensuring that Medicare is able to assess disparities is key to this goal. The ability to assess health disparities hinges on the quality of the underlying race and ethnicity data.

How OIG Did This Review

We analyzed the race and ethnicity data in Medicare's enrollment database, the only source of this information for all enrolled beneficiaries. These race and ethnicity data are derived from source data from the Social Security Administration (SSA) and the results of an algorithm that CMS applies to the source data. We assessed the accuracy of Medicare's enrollment race and ethnicity data for different groups by comparing them to self-reported data for a subset of beneficiaries who reside in nursing homes. Race and ethnicity data that are self-reported are considered the most accurate. We also assessed the adequacy of Medicare's data using the Federal standards for collecting race and ethnicity data as a benchmark.

What OIG Found

Medicare's enrollment race and ethnicity data are less accurate for some groups, particularly for beneficiaries identified as American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic. Data that are not accurate limit the ability to assess health disparities. Limited race and ethnicity categories and missing information contribute to inaccuracies in the enrollment data. Although the use of an algorithm improves the existing data to some extent, it falls short of self-reported data. Finally, Medicare's enrollment data on race and ethnicity are inconsistent with Federal data collection standards; these inconsistencies inhibit the work of identifying and improving health disparities within the Medicare population.

What OIG Recommends

Advancing health equity is a priority for CMS and the Department. Race and ethnicity data are foundational to identifying and understanding health disparities among Medicare beneficiaries and to assessing the effectiveness of efforts to reduce such disparities. It is critical that these data are accurate, complete, and comprehensive. Therefore, CMS must improve its race and ethnicity data; though a significant undertaking, the need for better data is pressing. Accordingly, we recommend that CMS: (1) develop its own source of race and ethnicity data, (2) use self-reported race and ethnicity information to improve data for current beneficiaries, (3) develop a process to ensure that the data are as standardized as possible, and (4) educate beneficiaries about CMS's efforts to improve the race and ethnicity information. CMS did not explicitly concur with the first recommendation and concurred with the other three recommendations.

Primer: Medicare's Race and Ethnicity Data

- Race, ethnicity, and other demographic information for Medicare beneficiaries are included in Medicare's enrollment data.
- Medicare's primary source of these race and ethnicity data is SSA. Historically, SSA collected race and ethnicity data on a voluntary basis when an individual applied for a Social Security card. SSA stopped routinely collecting these data in 1989 when individuals started receiving Social Security numbers when filing a birth certificate rather than applying for them separately, and now collects them only in certain, limited circumstances.
- Medicare improves the accuracy of the race and ethnicity data from SSA by imputing race and ethnicity data for some beneficiaries using the Research Triangle Institute (RTI) algorithm.²

Medicare's Race and Ethnicity Data



- Medicare also collects race and ethnicity data for beneficiaries in nursing homes. This information is self-reported and gathered during required assessments of beneficiaries' functional capabilities. It is included in the nursing home assessment data.³ Whereas the enrollment data include all Medicare beneficiaries, the nursing home assessment data include just Medicare beneficiaries in nursing homes.
- Current Federal standards for the collection of race and ethnicity data were developed in 1997 and 2011. Both standards recommend two questions, one for race and one for ethnicity; a comprehensive set of categories from which to choose; and allowing individuals to select all categories that apply. (See Appendix A for more information about the standardized categories.) Though these standards were developed more recently than Medicare's data were collected, they serve as a helpful benchmark for assessing the adequacy of Medicare's data.

→ Self-reported race and ethnicity information is considered the "gold standard" because it is the most accurate.*

→ Federal standards recommend self-reporting as the preferred method of collecting race and ethnicity data.

*Jarrín, Olga F.; Nyandegge, Abner N.; et al., "Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-Standard Self-Reported Race Collected During Routine Home Health Care Visits," *Medical Care*, Vol. 58, No. 1, January 2020, pp. e1–e8.

RESULTS

Medicare's enrollment data on race and ethnicity are less accurate for some groups, limiting the ability to assess and address health disparities

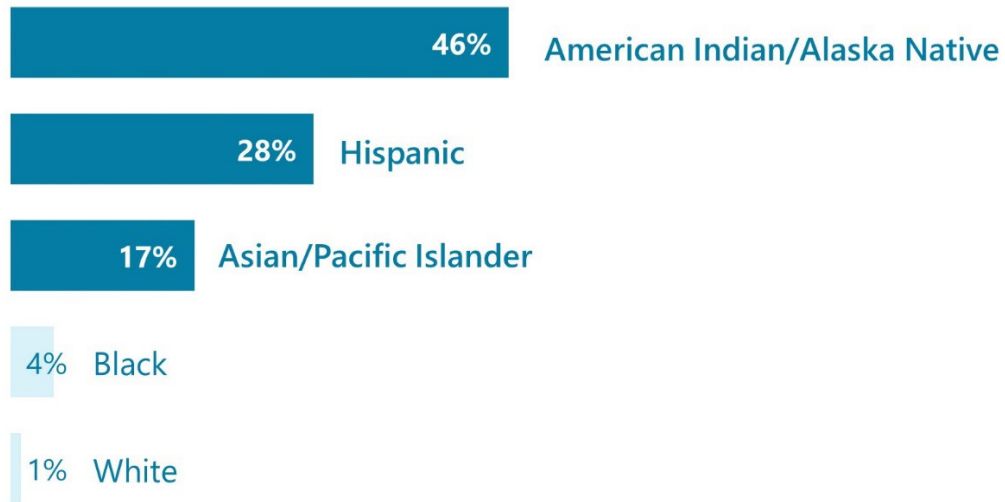
Medicare's enrollment data contain race and ethnicity information for about 66 million beneficiaries who were enrolled in Medicare at any point in 2020. However, this information is less accurate for some groups than for others when compared to self-reported data. Specifically, inaccuracies are more common for beneficiaries identified as American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic. Inaccuracies in race and ethnicity data have far-reaching consequences, affecting understanding of disparities in the prevalence, severity, and outcomes of diseases and conditions—including COVID-19—and in health care quality and access. As such, these inaccuracies limit the ability to effectively design and target efforts to address disparities in these areas and to measure the results of such efforts.

Nursing home assessment data contain self-reported race and ethnicity information for about 5 million beneficiaries who have been in nursing homes. Self-reported race and ethnicity data are considered the most reliable. When Medicare's enrollment data are compared to the self-reported race and ethnicity data, the enrollment data are much less accurate for some groups than for others.

Inaccuracies in the enrollment data reflect two errors. First, the enrollment data sometimes identify beneficiaries as a race and ethnicity with which they *do not* identify themselves on the nursing home assessment. For example, 28 percent of the beneficiaries identified as Hispanic in the enrollment data do not identify themselves as Hispanic on their nursing home assessments. The same error occurs for 46 percent of beneficiaries identified in the enrollment data as American Indian/Alaska Native and 17 percent of beneficiaries identified as Asian/Pacific Islander. (See Exhibit 1.)

- The nursing home assessment includes self-reported demographic data. Self-reported data are considered the most accurate.
- Residents report their race and ethnicity information on the assessment, but if the resident is not able to do so, a caregiver or clinician may respond.

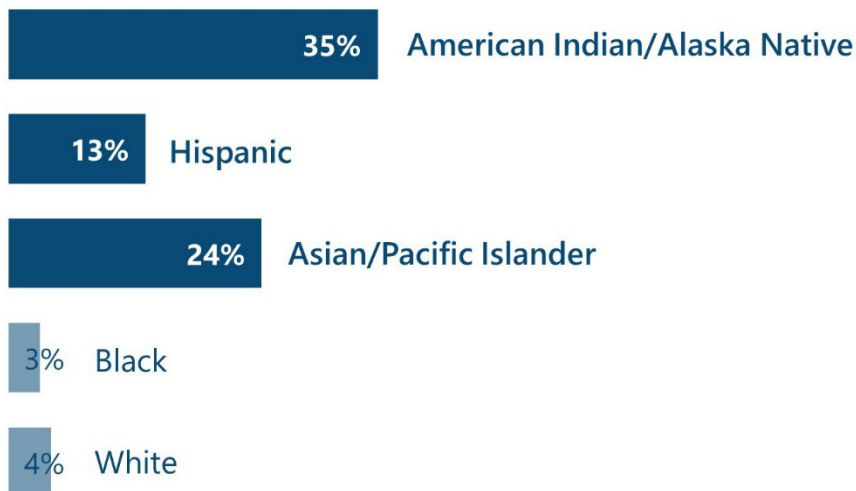
Exhibit 1: Percentage of Medicare Beneficiaries Identified in Medicare’s Enrollment Data as a Race and Ethnicity With Which They Do Not Identify Themselves on the Nursing Home Assessment



Source: OIG analysis of CMS data, 2021.

Second, the enrollment data do not capture the race and ethnicity with which these beneficiaries *do* identify. For example, 13 percent of beneficiaries who self-identified as Hispanic on their nursing home assessments are not identified that way in the enrollment data. The same error occurs for 35 percent of beneficiaries who self-identified as American Indian/Alaska Native on their nursing home assessments and for 24 percent of the beneficiaries who self-identified as Asian/Pacific Islander. In other words, almost one-quarter of the Asian/Pacific Islander population, as self-reported on nursing home assessments, are identified with another race and ethnicity or are missing race and ethnicity information in the enrollment data. (See Exhibit 2.)

Exhibit 2: Percentage of Medicare Beneficiaries Who Self-Identified as a Race and Ethnicity on Their Nursing Home Assessments but Are Not Identified That Way in Medicare’s Enrollment Data



Source: OIG analysis of CMS data, 2021.

Therefore, analyses to understand the unique health needs or outcomes of a racial or ethnic group will be based on faulty data if Medicare enrollment data are used. The data for each group will include some beneficiaries who are not in the group and will be missing other beneficiaries who should be in the group. The actual population of each group is different from the one identified by Medicare's enrollment data.

As a result, the usefulness of Medicare's enrollment data is diminished. The enrollment data can be a powerful source of information because they include all Medicare-enrolled beneficiaries. These data are used to evaluate the health needs and outcomes of the Medicare population and assess important issues.⁴ These issues could include racial and ethnic disparities in the long-term impacts of COVID-19, access to health care providers, or the quality of services provided. However, inaccurate race and ethnicity data can raise concerns that efforts to address disparities, to some extent, would be targeted at the wrong beneficiaries. Further, using inaccurate data to try to measure the effects of these efforts could produce misleading results.

Deficiencies in the source data contribute to inaccuracies in Medicare's enrollment race and ethnicity data

The primary source for Medicare's race and ethnicity information in the enrollment data is SSA. These source data have several limitations. Specifically, the data are limited to only a few race and ethnicity categories and are missing for more than 3 million beneficiaries.

The race and ethnicity categories are not comprehensive

The race and ethnicity categories used by SSA have historically been limited. Prior to 1980, beneficiaries could select from only three categories—White, Black, or Other—when applying for a Social Security card (individuals who did not complete the question were categorized as "Unknown").⁵ The SSA information for most beneficiaries enrolled in Medicare is limited to these categories.

Though SSA subsequently revised its categories, they remain limited in the data Medicare uses. Of note, they conflate race and ethnicity and beneficiaries are limited to just one category. These limitations mean that Medicare's data do not capture separate racial and ethnic identities (e.g., Black and Hispanic) or multiracial beneficiaries. Ultimately, these

Key Limitations of the Source Data

- Historically, the categories for race were limited to White, Black, or Other.
- The categories of American Indian/Alaska Native, Asian/Pacific Islander, and Hispanic were subsequently added as options in 1980.
- SSA stopped routinely collecting race and ethnicity data in 1989.

limitations impact Medicare's ability to accurately and meaningfully understand the health needs and outcomes of different populations.

Further, the limited categories contribute to the inaccuracies described above for the American Indian/Alaska Native, Asian/Pacific Islander, and Hispanic groups. They also likely explain the relatively low overall percentages of beneficiaries who are identified as Asian/Pacific Islander or Hispanic in the data from SSA. Only 2.5 percent are identified as Asian/Pacific Islander and 3 percent as Hispanic. For reference, a 2019 Profile of Older Americans estimates the over-65 population to be 5.1 percent Asian/Pacific Islander and 8 percent Hispanic.⁶

Data on race and ethnicity are lacking for more than 3 million beneficiaries and the completeness of the data will decrease over time

The data collected by SSA are lacking for 3.3 million Medicare beneficiaries. That is, 1.5 million beneficiaries are categorized as "Unknown" or have no race and ethnicity information in the SSA data. Another 1.8 million beneficiaries are included in the race and ethnicity category of "Other."

The completeness of the data from SSA upon which Medicare relies will only decrease with time. SSA stopped routinely collecting race and ethnicity information in 1989.⁷ As such, the SSA data will increasingly have missing or "Unknown" data. In fact, SSA no longer uses its data for race-related research because of its diminishing reliability.⁸

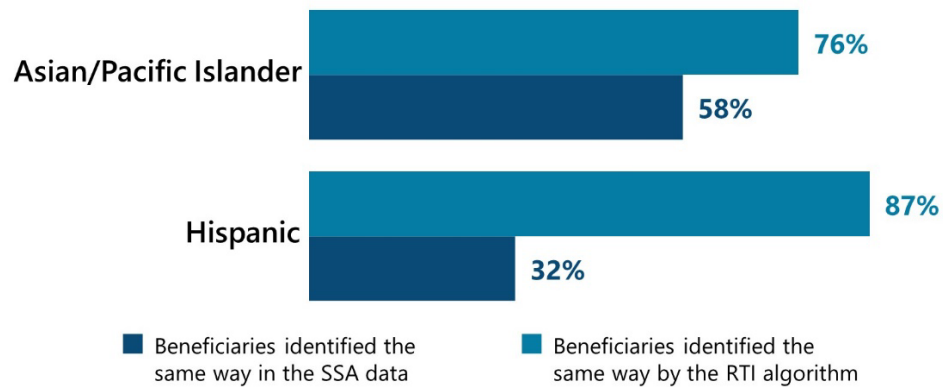
Using an algorithm improves the accuracy of Medicare's enrollment race and ethnicity data but falls short of self-reported data

To improve the quality of the information from SSA for the Asian/Pacific Islander and Hispanic groups, CMS applies an algorithm to that information in the Medicare enrollment data.⁹ This algorithm attempts to amend the race data from SSA by categorizing individuals based on name, geography, and requests for SSA and Medicare materials in Spanish. Though the RTI algorithm appears to improve the accuracy of race and ethnicity information for beneficiaries identified as Asian/Pacific Islander and Hispanic, the use of an algorithm has drawbacks.

The use of the algorithm increases the percentage of all beneficiaries identified as Asian/Pacific Islander from 2.5 percent in the SSA source data alone to 3.6 percent after the algorithm is applied and of beneficiaries identified as Hispanic from 3 percent to 8.7 percent. This means that Medicare is relying on the algorithm to supply race and ethnicity data for one-third of beneficiaries identified as Asian/Pacific Islander and two-thirds of beneficiaries identified as Hispanic. (See Appendix B for the distribution of the race and ethnicity data for all Medicare beneficiaries.)

Using the RTI algorithm improves the accuracy of the enrollment data for these two groups over using the data from SSA alone. For the subset of beneficiaries with nursing home assessments, the RTI algorithm improves upon the accuracy of the SSA data for those who self-identified as Asian/Pacific Islander by 18 percentage points. For those who self-identified as Hispanic, the RTI algorithm increases the accuracy by 55 percentage points. (See Exhibit 3.)

Exhibit 3: The RTI algorithm improves the accuracy of Medicare’s enrollment data for beneficiaries who self-identified as Hispanic or Asian/Pacific Islander on their nursing home assessments.



Source: OIG analysis of CMS data, 2021.

Despite improvements in accuracy, data derived by an algorithm relying on proxy information have more errors and are more inaccurate than self-reported data.¹⁰ (See the text box for more information about the RTI algorithm.)

Imputation of Race and Ethnicity Is Less Accurate Than Self-Reported Data

- The RTI algorithm relies on a list of last names that are frequently associated with Asian/Pacific Islander or Hispanic individuals.
 - As researchers and others have found, this practice does not accurately account for individuals who change their name after marriage, who are from colonized countries with names that closely align with the nationality of the colonizing nation, or who changed their name upon immigrating to the United States.*
- The RTI algorithm also uses geography for individuals living in Puerto Rico (Hispanic) or Hawaii (Asian/Pacific Islander).
- The RTI algorithm also factors in whether individuals requested SSA and Medicare materials in Spanish.
- Imputation may be appropriate for estimating race and ethnicity at the population level but is problematic when predicting an individual’s race and ethnicity.

*Jarrín, Olga F.; Nyandegge, Abner N.; et al., "Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-Standard Self-Reported Race Collected During Routine Home Health Care Visits," *Medical Care*, Vol. 58, No. 1, January 2020, pp. e1–e8. Roberts, Sam, "New Life in U.S. No Longer Means New Name," *New York Times*, August 25, 2010. Accessed at <https://www.nytimes.com/2010/08/26/nyregion/26names.html> on May 10, 2022.

Finally, even after applying the algorithm, Medicare lacks race and ethnicity information for 2 million beneficiaries. Specifically, 1.3 million beneficiaries still have “Unknown” or missing information and another 600,000 remain categorized as “Other.” It is also important to note that the RTI algorithm will not be able to compensate for the increasingly missing data from SSA because it is designed for the Asian/Pacific Islander and Hispanic groups only.

Medicare’s enrollment race and ethnicity data are inconsistent with Federal data collection standards

Federal standards—the 1997 Office of Management and Budget (OMB) standards and the 2011 standards developed by the Department of Health and Human Services (HHS) per the Patient Protection and Affordable Care Act—recommend that race and ethnicity data be collected in a comprehensive way. Comprehensive data collection is necessary to sufficiently understand health disparities and inform appropriate interventions and solutions. Medicare’s race and ethnicity data are inconsistent with key standards for the collection of such data, thereby limiting the usefulness of the data.

As noted above, Medicare’s enrollment data conflate race and ethnicity in one question and beneficiaries may select only one category. This conflicts with OMB and HHS standards, which recommend including two questions—one for race and one for ethnicity—and allowing individuals to select all categories that apply. The structure allows individuals to most accurately and fully represent both their racial and ethnic identities.

Key Federal Standards

- Separate questions for race and ethnicity
- Allowing respondents to select all categories that apply
- Comprehensive race and ethnicity categories

Further, the categories provided by Medicare’s enrollment data are not comprehensive. This conflicts with the HHS standards that expanded the categories of race and ethnicity data collection by adding granularity for the Hispanic and Asian/Pacific Islander groups. (See Appendix A for the categories.) Collecting these more granular data is crucial to knowledge about health disparities because a diverse racial or ethnic group may have very different health outcomes between its subgroups.¹¹ This type of specificity aids in more accurate analysis and more targeted interventions.

By expanding the categories on a large dataset like Medicare’s enrollment data to be consistent with the HHS standards, more robust analysis of race and ethnicity subcategories can be conducted. These categories are already employed on HHS surveys (e.g., Medicare Current Beneficiary Survey) and on other large-scale surveys (e.g., U.S. Census). Lack of consistency with the Federal standards in the Medicare enrollment data inhibits the work of identifying and reducing health disparities within the Medicare population.

CONCLUSION AND RECOMMENDATIONS

Comparisons to self-reported data show that Medicare's enrollment data on race and ethnicity are less accurate for some groups, particularly for beneficiaries identified as American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic. Data that are not accurate limit the ability to assess and address health disparities.

Deficiencies in the source data, such as categories that are not comprehensive and missing information, contribute to inaccuracies in Medicare's enrollment race and ethnicity data. In addition, using the RTI algorithm improves the accuracy of Medicare's enrollment race and ethnicity data but falls short of self-reported data. Finally, Medicare's enrollment race and ethnicity data are inconsistent with Federal data collection standards.

CMS has some work underway to improve the quality of the race and ethnicity data it collects. For example, it will incorporate race and ethnicity questions that are consistent with the HHS standards on the next versions of the post-acute assessments, such as the nursing home assessments. CMS has also presented strategies for improving the enrollment data in a Report to Congress mandated by the Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014.¹²

In the meantime, CMS seeks to expand the use of an alternative algorithm (Medicare Bayesian Improved Surname Geocoding) that it has found to be highly accurate.¹³ This algorithm relies on similar inputs as the RTI algorithm, as well as some additional data, such as Census-linked geocoding. We recognize that resolving the inaccuracies in the data identified in this report cannot happen overnight and we appreciate CMS's attention to this important issue. However, we note that this alternative algorithm also relies on the SSA source data and proxy information. Moreover, the alternative algorithm continues to use the same categories (limited to American Indian/Alaska Native, Asian/Pacific Islander, Black, Hispanic, and White) and to combine race and ethnicity.

Given the inaccuracies in Medicare's enrollment data and the priority of advancing health equity, CMS needs to improve its race and ethnicity data. Though this will be a significant undertaking, the need for better data is pressing. Accordingly, we recommend that CMS:

Develop its own source of race and ethnicity data

Given the limitations associated with the data from SSA and with imputation, CMS must transition away from the current process for obtaining race and ethnicity data and, instead, develop its own source of this information. This information should be voluntarily self-reported by beneficiaries and otherwise consistent with Federal standards.

For example, CMS could develop a way to collect race and ethnicity information from beneficiaries at Medicare enrollment, capturing information for newly enrolled beneficiaries, in coordination with SSA, as needed. CMS could also pursue alternative mechanisms for collection, such as online, by mail, and/or via the Parts C and D enrollment forms. Once collected, CMS should reconsider its need for imputation.

Use self-reported race and ethnicity information to improve data for current beneficiaries

As a more immediate step to improve the race and ethnicity data for current beneficiaries, CMS should incorporate the existing data from the nursing home and other post-acute care assessments, like the Outcome and Assessment Information Set (OASIS) assessment for home health care, into the enrollment data for current beneficiaries where available. These assessments allow beneficiaries to select more than one race or ethnicity, are ideally self-reported, and should be consistent with the HHS standards soon. Once established, new data from post-acute care assessments could be incorporated into the enrollment data on an ongoing basis.

Next, CMS should explore other ways of collecting self-reported data from current beneficiaries who do not have post-acute care assessments. There are many options for this. CMS could add race and ethnicity demographic questions to beneficiaries' online Medicare accounts. CMS could further collaborate with the Indian Health Service to improve the data for American Indian/Alaska Native beneficiaries. CMS could also conduct a targeted survey, similar to the postcard survey it used in the past, to reach beneficiaries not captured through other mechanisms.¹⁴ In choosing among these and other options, CMS should assess the number of beneficiaries likely impacted, the potential response rates, and the burdens on and benefits to beneficiaries and the Medicare program.

Develop a process to ensure that the data are as standardized as possible

CMS should develop a process to make the race and ethnicity data it collects as standardized as possible and able to be incorporated into the enrollment database. For example, CMS should standardize the phrasing of questions and response options pertaining to race and ethnicity and should use the Federal standards as a benchmark. Consistent and centralized data will allow for more comprehensive analyses and assessments of health disparities.

CMS should also develop a process for incorporating existing and new race and ethnicity data from different sources into the centralized enrollment database. It could start with the data from the nursing home and other post-acute care assessments. This process will need to address multiracial beneficiaries, which the enrollment database currently cannot accommodate, and beneficiaries who may report their information differently on multiple assessments.

Educate beneficiaries about CMS's efforts to improve the race and ethnicity information

CMS should educate beneficiaries about its efforts to improve the race and ethnicity data. Beneficiaries may be reluctant to provide this information to CMS if they do not understand how it will be used. CMS should develop and implement a plan to address these concerns that includes education and communication strategies. CMS should consider its own experiences and consult with others in determining appropriate and effective strategies for communicating with beneficiaries. These strategies could emphasize, for example, that responses do not affect Medicare coverage and include explanations for why the data are being collected at the point of collection (e.g., on an enrollment form, assessment, or Medicare account).

AGENCY COMMENTS AND OIG RESPONSE

CMS concurred with three of our recommendations and did not explicitly indicate whether it concurred with one recommendation. CMS emphasized its shared interest in self-reported and granular race and ethnicity data to better identify and address health disparities.

In response to our first recommendation to develop its own source of race and ethnicity data, CMS stated that it is committed to obtaining more accurate and comprehensive race and ethnicity data, consistent with its authorities and appropriation of funds under the law. CMS further noted that the President's Fiscal Year 2023 Budget includes a proposal to increase SSA sharing and collection of race and ethnicity data for individuals with Medicare. In the meantime, CMS stated that it is exploring opportunities to obtain more accurate and comprehensive race and ethnicity data. For example, CMS solicited comments on collecting race and ethnicity data on the Medicare Part C and D enrollment forms. In response, OIG emphasizes that the need for better data is pressing. We encourage CMS to continue pursuing these and other options and to seek statutory authority or request appropriations, as appropriate, to obtain more accurate and comprehensive race and ethnicity data.

CMS concurred with our second recommendation to use self-reported race and ethnicity information to improve data for current beneficiaries. CMS stated that it will continue exploring ways to obtain more accurate and comprehensive data for current beneficiaries and consider whether existing data, such as data obtained from nursing home or other post-acute care assessments, should be incorporated into the enrollment database.

CMS concurred with our third recommendation to develop a process to ensure that the data are as standardized as possible. CMS stated that it is exploring ways to standardize collected race and ethnicity data in a way that is compliant with up-to-date Federal standards.

CMS concurred with our fourth recommendation to educate beneficiaries about its efforts to improve the race and ethnicity information. CMS stated that it will explore opportunities to educate beneficiaries about its efforts to improve the collection of race and ethnicity data and how it relates to the agency's commitment to advancing health equity, expanding coverage, and improving health outcomes. CMS also noted it developed the Inventory of Resources for Standardized Demographic and Language Data Collection (updated March 2022) for stakeholders to improve the collection of standardized data, which can help them identify and address health disparities.

For the full text of CMS's comments, see Appendix C.

METHODOLOGY

Data Sources

We based this data brief on Medicare’s race and ethnicity data from the Master Beneficiary Summary File—referred to as the enrollment data. We included all beneficiaries enrolled in Medicare at any point in 2020. We also used data for any beneficiary who had at least one assessment in the Minimum Dataset (MDS)—referred to as the nursing home assessment data—with completed race and ethnicity data between January 1, 2018, and December 31, 2020. Lastly, we conducted interviews with CMS staff and other selected stakeholders to gain a better understanding of current data collection methods and strategies for improving the data.

Analysis of Accuracy of Medicare Race and Ethnicity Enrollment Data Compared to Self-Reported Nursing Home Assessment Data

To assess the accuracy of Medicare’s race and ethnicity data, we compared the RTI variable in the enrollment data with the self-reported nursing home assessment data in the MDS. For the purposes of this report, when this information did not match, we considered the enrollment data to be inaccurate, as self-reported data are considered the gold standard.¹⁵ First, we calculated the percentage of beneficiaries identified as a race and ethnicity in the enrollment data who did not self-identify themselves the same way on their nursing home assessments. Second, we calculated the percentage of beneficiaries who self-identified themselves as a race and ethnicity on their nursing home assessments but were not identified the same way in the enrollment data. For both measures, we calculated accuracy rates for each race and ethnicity category.¹⁶

To assess the relative accuracy of the RTI variable and the SSA source data, we identified beneficiaries who self-identified in the nursing home assessment data as either Asian/Pacific Islander or Hispanic. We then calculated the percentage that were identified as the same group in Medicare’s enrollment data for: (1) the SSA source data alone and (2) the data after the RTI algorithm was applied.

Analysis of Completeness of Medicare Race and Ethnicity Data

To assess the extent to which Medicare’s race and ethnicity data were missing, we calculated the number and percentage of beneficiaries by race and ethnicity category. We calculated the extent to which the information has values of “Other,” “Unknown,” or missing. We performed this calculation on both the data from SSA and on the data after CMS applies the RTI algorithm. We calculated the number and percentage of beneficiaries with incomplete race and ethnicity data for each variable by totaling the amount marked “Other” and the amount marked “Unknown” or missing.

Analysis of Federal Standards for Collection of Race and Ethnicity Data

We evaluated Medicare's enrollment data based on the 1997 OMB revised standards as well as the 2011 HHS standards per the Patient Protection and Affordable Care Act. Areas of analysis included race and ethnicity categories, separation or combination of race and ethnicity questions, and category selection process, such as single selection versus multiple selection. We also assessed the ways in which further alignment with these standards would improve Medicare's data. We used the OMB and HHS standards as a benchmark for assessing the adequacy and usefulness of Medicare's enrollment race and ethnicity data; this was not a compliance review.

Analysis of Interview Data

We conducted interviews with CMS staff and other stakeholders to further assess the history of the data collection process, as well as current and future efforts to adapt and improve race and ethnicity data collection from Medicare beneficiaries. Interviews covered strategies for and barriers to improving the data.

A Note About Terms

Some of the race and ethnicity categories have been described with different labels historically. Below is a complete list of the terms that have been used along with the one we use to refer to each category in this report.

- American Indian, North American Native, North American Indian, or Alaska Native (American Indian/Alaska Native);
- Asian, Asian American, Pacific Islander, or Native Hawaiian (Asian/Pacific Islander);
- Black or African American (Black); and
- Hispanic or Latino (Hispanic).

Limitations

Our analysis of the accuracy of the Medicare enrollment data is based on the subset of Medicare beneficiaries with a nursing home assessment. The racial and ethnic makeup of this subset differs from that of the overall Medicare population. The subset has a larger proportion of White and Black beneficiaries than the overall population.¹⁷

We did not assess all potential costs, barriers, or benefits related to collecting data directly from Medicare beneficiaries.

Standards

We conducted this study in accordance with the *Quality Standards for Inspection and Evaluation* issued by the Council of the Inspectors General on Integrity and Efficiency.

APPENDIX A

Federal Race and Ethnicity Data Collection Standards

Office of Management and Budget, 1997	HHS Implementation Guidance on Data Collection Standards, 2011
	Race
White	White
Black	Black or African American
American Indian or Alaska Native	American Indian or Alaska Native
Asian	Asian Indian
Native Hawaiian or Other Pacific Islander	Chinese
	Filipino
	Japanese
	Korean
	Vietnamese
	Other Asian
	Native Hawaiian
	Guamanian or Chamorro
	Samoan
	Other Pacific Islander
	Ethnicity
Not Hispanic or Latino	No, not of Hispanic, Latino/a, or Spanish origin
Hispanic or Latino	Yes, Mexican, Mexican American, Chicano/a
	Yes, Puerto Rican
	Yes, Cuban
	Yes, another Hispanic, Latino/a, or Spanish origin

Sources: Office of Management and Budget, *Federal Register* Vol. 62, No. 210, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity," October 10, 1997. <https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf>.

Department of Health and Human Services, "Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status," October 2011. <https://aspe.hhs.gov/system/files/pdf/76331/index.pdf>.

APPENDIX B

Distribution of Medicare Beneficiaries by Race and Ethnicity in Medicare's Enrollment Data

Race and Ethnicity	Percentage of Beneficiaries (SSA)	Percentage of Beneficiaries (RTI)	Number of Beneficiaries (RTI)
American Indian/Alaska Native	0.4%	0.4%	275,366
Asian/Pacific Islander	2.5%	3.6%	2,376,832
Black	10.8%	10.6%	6,992,296
Hispanic	3.0%	8.7%	5,730,381
White	78.3%	73.7%	48,583,154
Other	2.7%	1.0%	641,033
Unknown/Missing	2.3%	2.0%	1,340,795
Total	100%	100%	65,939,857

Source: OIG analysis of CMS data, 2021.



Administrator
Washington, DC 20201

DATE: May 3, 2022

TO: Gregory E. Demske
Acting Principal Deputy Inspector General
Office of Inspector General

FROM: Chiquita Brooks-LaSure *Chiquita LaSure*
Administrator
Centers for Medicare & Medicaid Services

SUBJECT: Office of Inspector General (OIG) Draft Report: Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities (OEI-02-21-00100)

The Centers for Medicare & Medicaid Services (CMS) appreciates the opportunity to review and comment on the Department of Health and Human Services (HHS) Office of Inspector General's (OIG) above-titled draft report.

CMS serves the public as a trusted partner and steward, dedicated to advancing health equity, expanding coverage, and improving health outcomes. CMS defines health equity as the attainment of the highest level of health for all people, where everyone has a fair and just opportunity to attain their optimal health regardless of race, ethnicity, disability, sexual orientation, gender identity, socioeconomic status, geography, preferred language, or other factors that affect access to care and health outcomes. CMS is working to advance health equity by designing, implementing, and operationalizing policies and programs that support health for all the people served by our programs, eliminating avoidable differences in health outcomes experienced by people who are disadvantaged or underserved, and providing the care and support that our enrollees need to thrive.

Central to these efforts is improving the collection, reporting, and analysis of CMS's health equity-related data. The Social Security Administration (SSA) is responsible for enrolling Medicare-eligible individuals into Medicare, and as such, is the primary source of race and ethnicity data for people with Medicare. However, SSA discontinued collecting these data for most people with Medicare born after 1989. The President's Fiscal Year (FY) 2023 Budget includes a proposal to increase SSA sharing and collection of race and ethnicity data for individuals with Medicare.¹ This administrative proposal would have SSA increase sharing of race and ethnicity data with CMS for current and prospective Medicare enrollees, and consider expanding collection of detailed data, *e.g.* at 2011 HHS Data Standards² or, when available, newer data standards. CMS would assist SSA by conducting appropriate research and user testing for collection of these data to ensure it is useful for the purposes of identifying and tracking disparities in healthcare treatment and outcomes by race and ethnicity.

¹ <https://www.hhs.gov/sites/default/files/fy-2023-budget-in-brief.pdf>

² <https://aspe.hhs.gov/reports/hhs-implementation-guidance-data-collection-standards-race-ethnicity-sex-primary-language-disability-0>

To augment the race and ethnicity data received from SSA, CMS has taken a variety of approaches to obtain or improve race and ethnicity information for people with Medicare including the use of indirect estimation methods (statistical imputation), conducting a postcard survey, and collaborating with the Indian Health Service. Most recently, in January 2022, CMS solicited comments regarding CMS' intention to pilot the collection of race and ethnicity data on the Medicare Part C and D enrollment forms.³

As part of our work to advance health equity, CMS has interest in identifying patterns of differences across many key process and care outcomes by sociodemographic characteristics, including race and ethnicity. To best characterize these differences, we agree that self-reported *and* granular race and ethnicity data are needed. Efforts to improve the collection of such data will continue to strengthen our work to advance health equity, expand coverage, and improve health outcomes. In addition, new data will allow us to validate the existing indirect estimation methods that CMS uses for race and ethnicity to ensure that we do not rely on methodologies that unintentionally create or exacerbate disparities.

We appreciate the OIG's work on this area and look forward to working collaboratively on this and other issues in the future.

The OIG's recommendations and CMS' responses are below.

OIG Recommendation

The OIG recommends that CMS develop its own source of race and ethnicity data.

CMS Response

CMS is committed to obtaining more accurate and comprehensive race and ethnicity data for people with Medicare, consistent with our authorities and appropriation of funds under the law. As stated above, the President's FY 2023 Budget includes a proposal to increase SSA sharing and collection of race and ethnicity data for individuals with Medicare. As SSA enrolls Medicare-eligible individuals into Medicare and shares that information with CMS, there is an opportunity to improve data collection through that mechanism.

In the meantime, CMS will continue to explore opportunities to obtain more accurate and comprehensive race and ethnicity data for people with Medicare consistent with our current authorities. For example, in January 2022, CMS solicited comments regarding our intention to pilot the collection of race and ethnicity data on the Medicare Part C and D enrollment forms. With the long-term goal of collecting race and ethnicity data from all people with Medicare, CMS plans to focus initial efforts on those who newly elect or change coverage in the Medicare Part C and D programs. The detailed race and ethnicity categories collected through the pilot demographic questions on the Medicare Part C and Part D enrollment forms would be compliant with the 2011 HHS Data Standards to provide granular information for plans and for CMS to understand the diversity of the Medicare population. If finalized, the data will be used to: (1) Explore the response rate to race and ethnicity questions as a whole and how it intersects with enrollee income and other demographics; (2) Conduct focus groups to understand how people who elect to not respond to the race and ethnicity questions perceive the addition of those questions on the form; (3) Continue to test CMS' race and ethnicity indirect estimation methods by adding additional race and ethnicity data to the data CMS already has; and (4) Determine the data necessary for sufficient samples sizes to conduct analyses of disaggregated race and ethnicity categories. In addition to the aforementioned uses, CMS will ultimately use this

³ <https://www.govinfo.gov/content/pkg/FR-2022-01-12/pdf/2022-00375.pdf>

information to: (1) Track enrollment, including tracking patterns in enrollment by race and ethnicity over time; (2) To identify, monitor, and develop effective and efficient strategies and incentives to reduce and eliminate health and health care inequities; (3) Validate existing race and ethnicity statistical imputation methods; and (4) Ensure that clinically appropriate and equitable care (in terms of payment, access, and quality) is consistently provided to all enrollees. The comment period for the proposal has ended, however, CMS is currently reviewing comments received.

OIG Recommendation

The OIG recommends that CMS use self-reported race and ethnicity information to improve data for current beneficiaries.

CMS Response

CMS concurs with this recommendation. CMS will continue to explore opportunities to obtain more accurate and comprehensive race and ethnicity data for people with Medicare consistent with the priorities and objectives of the Administration. CMS will consider whether any existing data, such as data obtained from nursing home or other post-acute care assessments, should be incorporated into the enrollment database.

OIG Recommendation

The OIG recommends that CMS develop a process to ensure that the data are as standardized as possible.

CMS Response

CMS concurs with this recommendation. CMS is exploring ways to standardize race and ethnicity data collected. For example, any new CMS data collection will be compliant with the 2011 HHS Data Standards or, when available, newer data standards. CMS will take the OIG's findings and recommendations in to account when determining any additional next steps.

OIG Recommendation

The OIG recommends that CMS educate beneficiaries about CMS's efforts to improve the race and ethnicity information.

CMS Response

CMS concurs with this recommendation. It is important that enrollees understand the value of this type of data and how the data will be utilized. It is equally important that enrollees understand that providing this type of data is voluntary and that providing such data is not required to get care or coverage and will not affect any payments. CMS will explore opportunities to educate its enrollees about CMS' efforts to improve the collection of race and ethnicity data and how it relates to the agency's commitment to advancing health equity, expanding coverage, and improving health outcomes through a variety of approaches and channels.

In addition, CMS believes that it is important to provide educational resources on this topic for other parties that may be looking for recommendations on how to best collect and use race and ethnicity data. For this purpose, CMS developed the Inventory of Resources for Standardized Demographic and Language Data Collection (updated March 2022).⁴ The inventory combines data collection best practices from reports, toolkits, webinars, and training tools into a single resource that can be used by health care organizations of all types. Stakeholders can use these resources to

⁴ <https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/Data-Collection-Resources.pdf>

improve the collection of standardized data, which can help them identify and address health disparities.

ACKNOWLEDGMENTS AND CONTACT

Acknowledgments

Rachel Bryan served as the team leader for this study. Other Office of Evaluation and Inspections staff from the New York regional office who conducted the study include Heather Koenig. Office of Evaluation and Inspections staff who provided support include Robert Gibbons and Michael Novello.

This report was prepared under the direction of Jodi Nudelman, Regional Inspector General for Evaluation and Inspections in the New York regional office, and Nancy Harrison and Meridith Seife, Deputy Regional Inspectors General.

Contact

To obtain additional information concerning this report, contact the Office of Public Affairs at Public.Affairs@oig.hhs.gov. OIG reports and other information can be found on the OIG website at oig.hhs.gov.

Office of Inspector General
Department of Health and Human Services
330 Independence Avenue, SW
Washington, DC 20201

ENDNOTES

¹ The Centers for Disease Control and Prevention, “Health Equity Considerations and Racial and Ethnic Minority Groups,” July 24, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html>.

² CMS contracted with RTI to create an algorithm that could be applied to the existing enrollment data from SSA to improve their accuracy. The RTI algorithm attempts to amend the race data for Hispanic and Asian/Pacific Islander beneficiaries by categorizing them based on name, geography, and requests made for SSA or Medicare materials in Spanish.

³ For beneficiaries in nursing homes, these data are collected on Minimum Dataset (MDS) assessments. Medicare also collects race and ethnicity data in other post-acute care settings, using different assessments, such as the Outcome and Assessment Information Set (OASIS) for home health care patients. Although the assessments are different, they all include the same race and ethnicity question. Beneficiaries report their race and ethnicity information on the assessments, but if the beneficiary is not able to do so, a caregiver or clinician may respond.

⁴ See, for example, CMS, “Preliminary Medicare COVID-19 Data Snapshot,” December 17, 2021, p. 6. <https://www.cms.gov/files/document/medicare-covid-19-data-snapshot-fact-sheet.pdf>.

⁵ Eicheldinger, Celia; Bonito, Arthur, “More Accurate Racial and Ethnic Codes for Medicare Administrative Data,” *Health Care Financing Review*, Vol. 29, No. 3, Spring 2008. <https://pubmed.ncbi.nlm.nih.gov/18567241/>.

⁶ The Administration for Community Living, “2019 Profile of Older Americans,” May 2020. <https://acl.gov/sites/default/files/Aging%20and%20Disability%20in%20America/2019ProfileOlderAmericans508.pdf>. This report relied on self-reported race and ethnicity data from sources such as the U.S. Census Bureau.

⁷ Scott, Charles G., “Identifying the Race or Ethnicity of SSI Recipients,” *Social Security Bulletin*, Vol. 62, No. 4, 1999.

⁸ SSA instead uses survey data. See Martin, Patricia P., “Why Researchers Now Rely on Surveys for Race Data on OASDI and SSI Programs: A Comparison of Four Major Surveys,” Social Security Office of Retirement and Disability Policy, Research, and Statistics Note No. 2016-01. <https://www.ssa.gov/policy/docs/rsnotes/rsn2016-01.html>.

⁹ In addition to the algorithm, CMS has taken two other actions to improve the SSA source data. In 1997, CMS sent a postcard survey to individuals with Hispanic surnames and geographical locations to more accurately identify Hispanic beneficiaries, an underrepresented category in the SSA data. Eicheldinger, Celia; Bonito, Arthur, “More Accurate Racial and Ethnic Codes for Medicare Administrative Data,” *Health Care Financing Review*, Vol. 29, No. 3, Spring 2008, p. 2. <https://pubmed.ncbi.nlm.nih.gov/18567241/>.

And, in 1999, CMS began collaborating with the Indian Health Service to better capture American Indian and Alaska Native individuals in the enrollment data. CMS, “Report to Congress: Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014 Strategic Plan for Accessing Race and Ethnicity Data,” January 5, 2017, p. 9. <https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/Research-Reports-2017-Report-to-Congress-IMPACT-ACT-of-2014.pdf>.

According to our analysis, the postcard survey and Indian Health Service adjustments improved race and ethnicity data for 0.6 percent of current Medicare beneficiaries.

¹⁰ Jarrín, Olga F.; Nyandegge, Abner N.; et al., “Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-Standard Self-Reported Race Collected During Routine Home Health Care Visits,” *Medical Care*, Vol. 58, No. 1, January 2020, pp. e1–e8.

¹¹ Gordan, Nancy, et al., “Aggregation of Asian-American subgroups masks meaningful differences in health and health risks among Asian ethnicities: an electronic health record based cohort study,” November, 25, 2019. <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7683-3>.

¹² CMS, "Report to Congress: Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014 Strategic Plan for Accessing Race and Ethnicity Data," January 5, 2017. <https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/Research-Reports-2017-Report-to-Congress-IMPACT-ACT-of-2014.pdf>.

¹³ CMS, "Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2022 Rates; Quality Programs and Medicare Promoting Interoperability Program Requirements for Eligible Hospitals and Critical Access Hospitals; Changes to Medicaid Provider Enrollment; and Changes to the Medicare Shared Savings Program," 86 Fed. Reg. 44774, 45264 (Aug. 13, 2021). <https://www.govinfo.gov/content/pkg/FR-2021-08-13/pdf/2021-16519.pdf>.

¹⁴ Eicheldinger, Celia; Bonito, Arthur, "More Accurate Racial and Ethnic Codes for Medicare Administrative Data," *Health Care Financing Review*, Vol. 29, No. 3, Spring 2008, p. 2. <https://pubmed.ncbi.nlm.nih.gov/18567241/>.

¹⁵ Self-reported data are used as the "gold-standard" to assess the accuracy of other data sources. Eicheldinger, Celia; Bonito, Arthur, "More Accurate Racial and Ethnic Codes for Medicare Administrative Data," *Health Care Financing Review*, Vol. 29, No. 3, Spring 2008. <https://pubmed.ncbi.nlm.nih.gov/18567241/>.

¹⁶ For beneficiaries with more than one nursing home assessment during our timeframe, we used the race and ethnicity data from the most recent one.

¹⁷ The overall Medicare population is 74 percent White and 11 percent Black, while the subset with a nursing home assessment is 78 percent White and 12 percent Black, according to the RTI variable.