

U.S. Department of Health and Human Services
Office of Inspector General



***A Resource Guide for Using
Medicare's Enrollment Race
and Ethnicity Data***

Ann Maxwell
Deputy Inspector General
for Evaluation and Inspections
June 2023, OEI-02-21-00101





A Resource Guide for Using Medicare's Enrollment Race and Ethnicity Data

Key Takeaway

The race and ethnicity data in Medicare's enrollment data have limitations, but can still be used in important health equity work if those limitations are recognized and accounted for during analysis.

Why did OIG create this resource guide?

Medicare is an essential part of the Nation's health care system, with 66 million people enrolled. The COVID-19 pandemic brought persistent disparities in health care access and outcomes to the forefront, including in the Medicare program. The Office of Inspector General (OIG) and the Centers for Medicare & Medicaid Services (CMS) have made advancing health equity a top priority. In order to address health disparities, it is important to assess them using accurate, complete, and comprehensive data. The results of these analyses can be used to tailor interventions aimed at improving disparities. The data can then be used to evaluate the efficacy of these interventions. Ultimately, success in advancing health equity hinges on a thorough understanding of the underlying data.

In June 2022, OIG issued a data brief, *Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities* (OEI-02-21-00100), analyzing the quality of the race and ethnicity data for people enrolled in Medicare. That data brief makes constructive recommendations to CMS for improving the data.¹

As CMS works to implement these recommendations, improvements are expected to take several years to yield better data for health disparities research.² Until then, the existing data remain a vital source for understanding one of the largest Federal programs and conducting health equity work.

What is the resource guide?

This resource guide is a tool that provides practical information about the race and ethnicity data for people enrolled in Medicare. Specifically, using information derived from the prior data brief, it provides a concise, user-friendly explanation of the origins and limitations of Medicare's enrollment race and ethnicity data and offers considerations for the use of these data.

This guide is for public and private sector users of Medicare data, such as researchers, managed care organizations, Congress, and others, involved in health equity work.

RESOURCE GUIDE

This guide includes the following sections: (1) overview of Medicare's race and ethnicity data; (2) origins of Medicare's enrollment race and ethnicity data; (3) understanding the limitations of the data; (4) developing findings based on these data; and (5) available self-reported race and ethnicity data for some people enrolled in Medicare.

Overview of Medicare's race and ethnicity data

Race, ethnicity, and other demographic information for people enrolled in Medicare are included in Medicare's enrollment data. The enrollment data are the **only source of race and ethnicity data for all people enrolled in Medicare**.

- These enrollment data are featured in the Master Beneficiary Summary File (MBSF), the Chronic Condition Warehouse (CCW), and the Integrated Data Repository (IDR).

Origins of Medicare's enrollment race and ethnicity data

Knowing more about where Medicare's data come from may be useful for developing effective analyses and findings; understanding their limitations; and communicating them. There are two race and ethnicity variables available from Medicare's enrollment data:

BENE_RACE_CD: This is the original variable for race and ethnicity, and its primary source is the Social Security Administration (SSA).

- Prior to 1980, people applying for Social Security cards could select from only three categories—White, Black, or Other—when voluntarily providing their race and ethnicity information.
- In 1980, the categories were expanded to American Indian/Alaska Native, Asian/Pacific Islander, Black, Hispanic, White, and Other.
- SSA stopped routinely collecting these data in 1989, though information for most current Medicare beneficiaries was collected prior to this date.³

RTI_RACE_CD: To improve information for the Asian/Pacific Islander and Hispanic groups, who were not captured in the original response categories, CMS applies an algorithm developed by the Research Triangle Institute (RTI) to the BENE_RACE_CD data.⁴ This creates the RTI_RACE_CD variable.

- The RTI_RACE_CD variable includes all of the BENE_RACE_CD data. CMS then updates these data for certain individuals by using the RTI algorithm, which categorizes individuals as Asian/Pacific Islander or Hispanic on the basis of name, geography, and requests for SSA and Medicare materials in Spanish.

Of the two variables, the **RTI_RACE_CD data are more accurate and complete** than the BENE_RACE_CD data and are what we use in our analyses.⁵

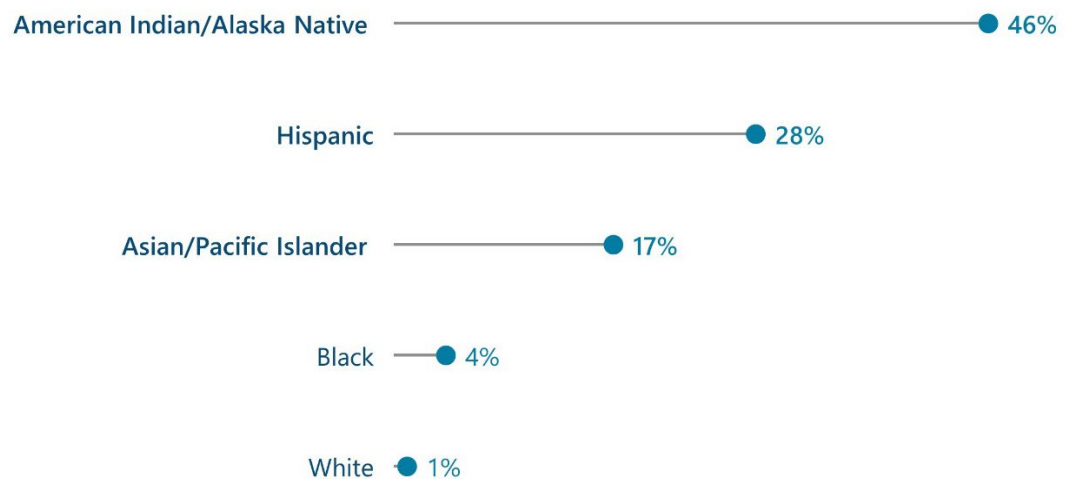
Understanding the limitations of the data

Although the RTI_RACE_CD data are more accurate and complete than the BENE_RACE_CD data for the Asian/Pacific Islander and Hispanic groups, there are some limitations associated with the RTI_RACE_CD data that warrant consideration when conducting analyses.

Accuracy

The RTI_RACE_CD data are not as accurate as self-reported data, which are considered the “gold standard.”⁶ When the RTI_RACE_CD data were compared to self-reported data for a subset of Medicare enrollees with nursing home assessments, the RTI_RACE_CD data were **less accurate for people identified as American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic.**⁷ For example, 28 percent of the people identified as Hispanic in the RTI_RACE_CD data do not identify themselves as Hispanic on their nursing home assessments. Also note that, as designed, the RTI algorithm does not address accuracy concerns related to the American Indian/Alaska Native group. See Exhibit 1.

Exhibit 1: Percentage of Medicare Enrollees Identified in Medicare’s RTI_RACE_CD 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.

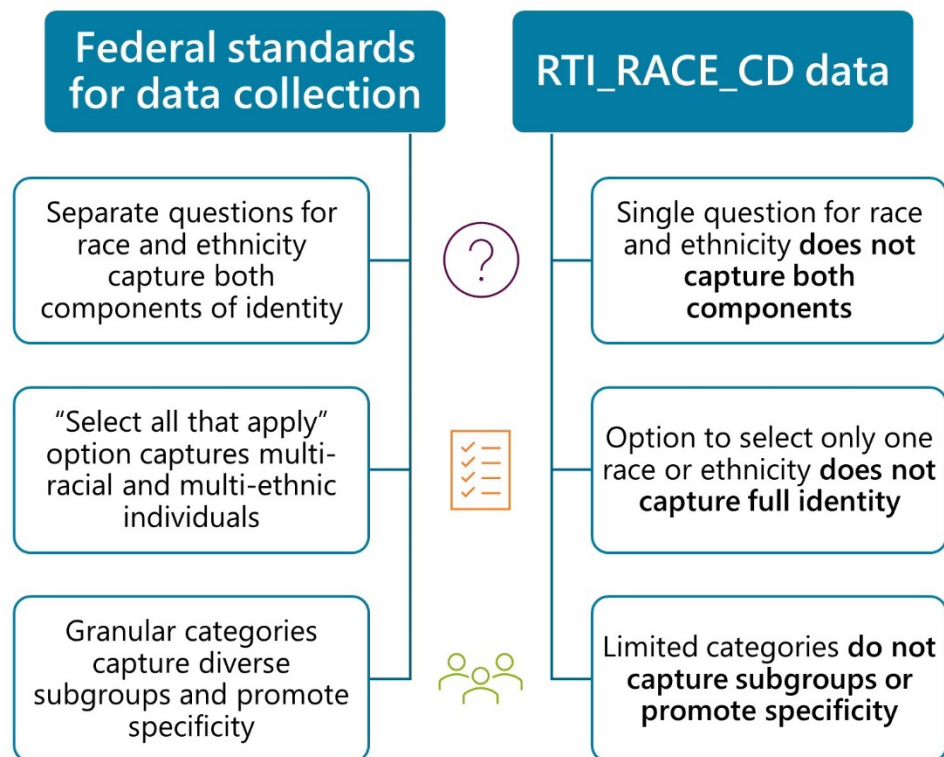
Completeness

Of approximately 66 million people enrolled in Medicare in 2020, a total of **2 million were lacking race and ethnicity information** in the RTI_RACE_CD data, with 1.3 million categorized as “Unknown” or missing, and 600,000 categorized as “Other.” Because of the way the SSA source data were collected, these people are more likely to be American Indian/Alaska Native, Asian/Pacific Islander, and Hispanic.

Collection methods

The RTI_RACE_CD data are **inconsistent with Federal standards** for the collection of race and ethnicity data, which are designed to ensure consistent data collection and appropriately detailed data.⁸ For example, Federal standards recommend requesting race and ethnicity via two separate questions to capture both racial and ethnic identities and multi-racial individuals. However, the RTI_RACE_CD data combine race and ethnicity and limit each person to only one race or ethnicity. See Exhibit 2.

Exhibit 2: Notable Differences Between Federal Standards for the Collection of Race and Ethnicity Data and Medicare’s RTI_RACE_CD Data



Developing findings based on these data

Important conclusions can still be drawn using the RTI_RACE_CD data despite their limitations.

Users of the RTI_RACE_CD data should consider the limitations of the data when conducting their analyses and communicate decisions they have made based on these limitations. For example, users of the data may want to include statements about the overall accuracy of the data; their decisions about combining data or not reporting on certain data; or background information about how the data were collected.

Taking these considerations into account, OIG has used the RTI_RACE_CD data in studies assessing health disparities. For examples of OIG's findings using these data in our reports, see Exhibit 3 below.⁹ Note that OIG chose not to report findings based on the RTI_RACE_CD data for the American Indian/Alaska Native group in these three instances because of concerns about accuracy. CMS has also used the RTI_RACE_CD data in its own analyses of COVID-19 among Medicare enrollees.¹⁰

Exhibit 3: Select OIG Findings on Health Disparities Using the RTI_RACE_CD Data

- Asian/Pacific Islander, Hispanic, and Black Medicare enrollees were less likely to receive medication than White enrollees to treat their opioid use disorder.
- Hispanic people enrolled in Medicare were more likely than others to use telehealth during the COVID-19 pandemic.
- Black and Hispanic people enrolled in Medicare were disproportionately hospitalized with COVID-19.

In addition, below is an **example of a statement** about its race and ethnicity data analysis that OIG included in one of its studies.

Note that Medicare data combine race and ethnicity and allow for only one response. Furthermore, we did not include information on American Indian/Alaska Native [enrollees] because of limitations. This analysis uses the race and ethnicity information from Medicare's enrollment data, which is based on data collected from the Social Security Administration and an algorithm. Although this information is currently the best available for the entire Medicare [enrollee] population, comparisons to self-reported data (available in certain, limited circumstances) show that race and ethnicity is still misclassified for some [enrollees]. In particular, Medicare [enrollees] with a race and ethnicity of American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic are more likely to be misclassified.¹¹

Available self-reported race and ethnicity data for some people enrolled in Medicare

Self-reported race and ethnicity data are available for subsets of the Medicare population and should be considered for use if appropriate to the study topic. As noted earlier, self-reported race and ethnicity data are considered the gold standard.

- Self-reported race and ethnicity data are **collected on post-acute care assessments** such as the Minimum Data Set (MDS) used in certified nursing facilities and the Outcome and Assessment Information Set (OASIS) used in home health care. These assessments collect and report data on quality of care for individuals receiving post-acute care services such as nursing home care and home health care.
- OIG has used these data to look at health disparities. For example, an OIG study of COVID-19 in nursing homes used MDS self-reported data. The study found that about half of Black, Hispanic, and Asian residents in nursing homes had or likely had COVID-19 in 2020 while 41 percent of White residents did.¹²
- The Medicare Current Beneficiary Survey is another source of self-reported race and ethnicity data from a random sample of enrollees.

METHODOLOGY

This resource guide is based on our recent data brief, *Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities* (OEI-02-21-00100).¹³ The resource guide highlights aspects of our findings and discusses the implications of those findings that are most relevant to users of Medicare's race and ethnicity data. See the methodology section of the recent data brief for information about how we assessed the quality of the race and ethnicity 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 resource guide.

- 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

This guide is meant to serve as a practical resource for those who may use the Medicare enrollment race and ethnicity data; it is not meant to be a prescriptive document about the way the data must be used to conduct analyses.

Our analysis of the accuracy of the Medicare enrollment data is based on the subset of Medicare enrollees 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.¹⁴

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.

Notice

This resource guide was prepared as a technical resource and is not intended to be used to determine compliance with any laws, regulations, or other guidance. In no event shall OIG or its employees, agents, or staff be liable for any claim, damages, or liability, whether in an action of contract, tort, or otherwise, and including direct, indirect, incidental, special, exemplary, or consequential damages, however caused, and on any theory of liability, arising in any way out of the use of this resource guide, even if advised of the possibility of such damage.

ACKNOWLEDGMENTS AND CONTACT

Acknowledgments

Rachel Bryan served as the team leader for this study. Others in the Office of Evaluation and Inspections who conducted the study include Heather Koenig. Office of Evaluation and Inspections headquarters 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 Meredith 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
U.S. Department of Health and Human Services
330 Independence Avenue, SW
Washington, DC 20201

ABOUT THE OFFICE OF INSPECTOR GENERAL

The mission of the Office of Inspector General (OIG), as mandated by Public Law 95-452, as amended, is to protect the integrity of the Department of Health and Human Services (HHS) programs, as well as the health and welfare of beneficiaries served by those programs. This statutory mission is carried out through a nationwide network of audits, investigations, and inspections conducted by the following operating components:

The Office of Audit Services (OAS) provides auditing services for HHS, either by conducting audits with its own audit resources or by overseeing audit work done by others. Audits examine the performance of HHS programs and/or its grantees and contractors in carrying out their respective responsibilities and are intended to provide independent assessments of HHS programs and operations. These audits help reduce waste, abuse, and mismanagement and promote economy and efficiency throughout HHS.

The Office of Evaluation and Inspections (OEI) conducts national evaluations to provide HHS, Congress, and the public with timely, useful, and reliable information on significant issues. These evaluations focus on preventing fraud, waste, or abuse and promoting economy, efficiency, and effectiveness of departmental programs. To promote impact, OEI reports also present practical recommendations for improving program operations.

The Office of Investigations (OI) conducts criminal, civil, and administrative investigations of fraud and misconduct related to HHS programs, operations, and beneficiaries. With investigators working in all 50 States and the District of Columbia, OI utilizes its resources by actively coordinating with the Department of Justice and other Federal, State, and local law enforcement authorities. The investigative efforts of OI often lead to criminal convictions, administrative sanctions, and/or civil monetary penalties.

The Office of Counsel to the Inspector General (OCIG) provides general legal services to OIG, rendering advice and opinions on HHS programs and operations and providing all legal support for OIG's internal operations. OCIG represents OIG in all civil and administrative fraud and abuse cases involving HHS programs, including False Claims Act, program exclusion, and civil monetary penalty cases. In connection with these cases, OCIG also negotiates and monitors corporate integrity agreements. OCIG renders advisory opinions, issues compliance program guidance, publishes fraud alerts, and provides other guidance to the health care industry concerning the anti-kickback statute and other OIG enforcement authorities.

ENDNOTES

¹ In the data brief, OIG recommended 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.

² CMS has not yet implemented the data brief's four recommendations, but CMS has added race and ethnicity questions to the 2023 Medicare Part C/D enrollment forms and plans to collaborate with the Social Security Administration about collecting these data for Medicare Part A/B enrollees.

CMS, *The Path Forward: Improving Data to Advance Health Equity Solutions*, November 2022.

<https://www.cms.gov/files/document/path-forwardhe-data-paper.pdf>.

³ In 1989, individuals started receiving Social Security numbers when their birth certificates were filed rather than applying for them separately. Now, race and ethnicity information are only collected in certain limited circumstances.

⁴ CMS has taken two other actions to add to the SSA source data in BENE_RACE_CD. In 1997, CMS sent a postcard survey to individuals with Hispanic surnames and geographical locations to more accurately identify Hispanic enrollees, 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/>.

Also, 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 Medicare enrollees in 2020.

⁵ CMS has developed an alternative algorithm called "Medicare Bayesian Improved Surname Geocoding" or "MBISG" that is not yet widely available for CMS's enrollment data. This algorithm relies on similar inputs as the RTI algorithm, as well as some additional data, such as Census-linked geocoding. We note that this alternative algorithm also relies on the SSA source data and proxy information. Moreover, the alternative algorithm continues to use limited categories with the addition of Multiracial (American Indian/Alaska Native, Asian/Pacific Islander, Black, Hispanic, White, and Multiracial) and to combine race and ethnicity.

⁶ 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/>.

⁷ OIG, *Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities* (OEI-02-21-00100), June 2022. <https://oig.hhs.gov/oei/reports/OEI-02-21-00100.pdf>.

⁸ Federal standards include those published by the Office of Management and Budget in 1997 and by the Department of Health and Human Services, per the Patient Protection and Affordable Care Act, in 2011.

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 30, 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>.

⁹ OIG, *Many Medicare Beneficiaries Are Not Receiving Medication to Treat Their Opioid Use Disorder* (OEI-02-20-00390), December 2021. <https://oig.hhs.gov/oei/reports/OEI-02-20-00390.pdf>.

OIG, *Certain Medicare Beneficiaries, Such as Urban and Hispanic Beneficiaries, Were More Likely Than Others To Use Telehealth During the First Year of the COVID-19 Pandemic* (OEI-02-20-00522), September 2022. <https://oig.hhs.gov/oei/reports/OEI-02-20-00522.pdf>.

OIG, *Medicare Beneficiaries Hospitalized With COVID-19 Experienced a Wide Range of Serious, Complex Conditions* (OEI-02-20-00410), September 2021. <https://oig.hhs.gov/oei/reports/OEI-02-20-00410.pdf>.

¹⁰ 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>.

¹¹ OIG, *Many Medicare Beneficiaries Are Not Receiving Medication to Treat Their Opioid Use Disorder* (OEI-02-20-00390), December 2021. <https://oig.hhs.gov/oei/reports/OEI-02-20-00390.pdf>.

¹² OIG, *COVID-19 Had a Devastating Impact on Medicare Beneficiaries in Nursing Homes During 2020* (OEI-02-20-00490), June 2021. <https://oig.hhs.gov/oei/reports/OEI-02-20-00490.pdf>.

¹³ OIG, *Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities* (OEI-02-21-00100), June 2022. <https://oig.hhs.gov/oei/reports/OEI-02-21-00100.pdf>.

¹⁴ 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.