

RESEARCH ARTICLE

Assessing the validity of race and ethnicity coding in administrative Medicare data for reporting outcomes among Medicare advantage beneficiaries from 2015 to 2017

Andrew W. Huang MS  | David J. Meyers PhD 

Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, Rhode Island, USA

Correspondence

Andrew W. Huang, MS, Department of Health Services, Policy and Practice, Brown University School of Public Health, 121 S Main Street, Providence, RI 02912, USA.
Email: andrew_huang2@brown.edu

Funding information

National Institutes of Health, Grant/Award Number: 1R21MD016147

Abstract

Objective: To assess the validity of race/ethnicity coding in Medicare data and whether misclassification errors lead to biased outcome reporting by race/ethnicity among Medicare Advantage beneficiaries.

Data Sources and Study Setting: In this national study of Medicare Advantage beneficiaries, we analyzed individual-level data from the Health Outcomes Survey (HOS) and the Consumer Assessment of Healthcare Providers and Systems (CAHPS), race/ethnicity codes from the Medicare Master Beneficiary Summary File (MBSF), and outcomes from the Medicare Provider Analysis and Review (MedPAR) files from 2015 to 2017.

Study Design: We used self-reported beneficiary race/ethnicity to validate the Medicare Enrollment Database (EDB) and Research Triangle Institute (RTI) race/ethnicity codes. We measured the sensitivity, specificity, and positive and negative predictive values of the Medicare EDB and RTI codes compared to self-report. For outcomes, we compared annualized hospital admission, 30-day, and 90-day readmission rates.

Data Collection/Extraction Methods: Data for Medicare Advantage beneficiaries who completed either the HOS or CAHPS survey were linked to MBSF and MedPAR files. Validity was assessed for both self-reported multiracial and single-race beneficiaries.

Principal Findings: For beneficiaries enrolled in Medicare Advantage, the EDB and RTI race/ethnicity codes have high validity for identifying non-Hispanic White or Black beneficiaries, but lower sensitivity for beneficiaries self-reported Hispanic any race (EDB: 28.3%, RTI: 85.9%) or non-Hispanic Asian American or Native Hawaiian Pacific Islander (EDB: 56.1%, RTI: 72.1%). Only 8.7% of beneficiaries self-reported non-Hispanic American Indian Alaska Native are correctly identified by either Medicare code, resulting in underreported annualized hospitalization rates (EDB: 31.5%, RTI: 31.6% vs. self-report: 34.6%). We find variation in 30-day readmission rates for Hispanic beneficiaries across race categories, which is not measured by Medicare race/ethnicity coding.

Conclusions: Current Medicare race/ethnicity codes misclassify and bias outcomes for non-Hispanic AIAN beneficiaries, who are more likely to select multiple racial identities. Revisions to race/ethnicity categories are needed to better represent multiracial/ethnic identities among Medicare Advantage beneficiaries.

KEYWORDS

ethnicity, health inequities, Medicare, minority health, racial groups, validation study

What is known on this topic

- Previous validation studies of single-race beneficiaries show that Medicare race/ethnicity codes perform well at identifying non-Hispanic White and non-Hispanic Black beneficiaries.
- However, Medicare race/ethnicity codes perform worse for beneficiaries identifying as Hispanic any race, non-Hispanic Asian American or Native Hawaiian Pacific Islander (AANHPI), or non-Hispanic American Indian and Alaska Native (AIAN).

What this study adds

- For Medicare Advantage, which enrolls a greater proportion of racial/ethnic minorities than traditional Medicare, we find that race/ethnicity codes correctly identify only 8.7% of non-Hispanic AIAN beneficiaries in our study.
- Beneficiaries who self-report multiple race categories, including 85% of the non-Hispanic AIAN beneficiaries in our study, are more likely to be misclassified by and have underreported hospitalization and readmission outcomes.
- Current Medicare race/ethnicity coding combines Hispanic ethnicity with race categories and exclude multiracial identity, which masks variation in hospital admission and readmission outcomes among self-reported Hispanic beneficiaries by race.

1 | INTRODUCTION

Recent Centers for Medicare and Medicaid Services (CMS) initiatives on health equity have centered around goals to reduce disparities and inequities in health outcomes among racial and ethnic minorities.^{1,2} To achieve these goals, careful and rigorous measurement of individual-level demographic and social needs data is critical, including the collection of beneficiary race and ethnicity.^{3,4} However, it is unclear to what extent the current datasets used by policy makers and researchers adequately capture beneficiary race/ethnicity, particularly for those who identify with multiple races or ethnicities.

Demographic information collected in the Medicare administrative records is a valuable source of data for measuring racial and ethnic health disparities. Beneficiary race and ethnicity is historically sourced from Social Security Administration enrollment and stored as the race/ethnicity variable in the Medicare Enrollment Database (EDB).^{5,6} However, the Social Security SS-5 application form contained only three categories for race and ethnicity (“White,” “Black,” or “Other”), and it was not until 1980 that the SS-5 form expanded from 3 to 5 categories (“non-Hispanic White,” “non-Hispanic Black,” “Hispanic,” “North American Indian or Alaskan Native,” or “Asian, Asian-American or Pacific Islander”), with the Health Care Financing Administration (now CMS) officially adopting these changes to the EDB code in 1994.^{5,6} At the time of this change, survey questionnaires were mailed to request race and ethnicity information from beneficiaries previously classified as “Other” or “Unknown,” but there have otherwise been few opportunities for beneficiaries to update their race and ethnicity after Medicare enrollment.^{5,6} Several validation studies have raised questions about the accuracy of these

updates and found that the EDB code has undercounted beneficiaries identifying as “Hispanic,” “North American Indian or Alaskan Native,” or “Asian, Asian-American or Pacific Islander.”⁵⁻⁸

To supplement the EDB code, CMS applies an imputation algorithm developed by the Research Triangle Institute (RTI) to categorize Medicare beneficiaries based on their name, geography, and requests for language materials, with predictions incorporated into the Medicare administrative database as a single-race/ethnicity code known as the “RTI race code.”^{9,10} The RTI race code has been used by health services researchers to measure racial and ethnic disparities in outcomes and quality of care among Medicare beneficiaries^{11,12} and has also been used by CMS for measuring racial and ethnic disparities in hospitalization outcomes reporting, including recent COVID-19 Hospitalization Trend Reports.¹³

However, any improvement in the accuracy of the RTI race code over the EDB code should be interpreted with caution. Unlike the EDB code, which represents self-reported race and ethnicity sourced from Social Security enrollment forms, the RTI race code represents a derived prediction of race and ethnicity based on an imputation algorithm. Any misclassification of race and ethnicity by the EDB code should be considered an error due to response bias. In contrast, any misclassification of race and ethnicity by the RTI imputation algorithm could be representative of error due to both response bias and prediction model bias. If the RTI race code disproportionately misclassifies beneficiaries from certain race and ethnicity groups, there is risk of additional algorithmic biases being introduced when reporting any outcomes of interest stratified by predicted race/ethnicity.^{14,15} Instead, the use of self-reported race and ethnicity should be considered the gold-standard for identifying beneficiary race and ethnicity.^{16,17}

Recent studies have assessed the validity of EDB and RTI race/ethnicity coding in Medicare data using 2015 home health¹⁸ and 2018–2020 nursing home¹⁹ assessments; however, the populations studied are specific to beneficiaries receiving postacute care, and it is unclear how consistently beneficiary race and ethnicity is collected during assessment visits compared to self-report. Another study assessed the validity of the RTI race code using self-reported race and ethnicity information from 2011 to 2015 American Community Survey responses linked to Medicare enrollment but did not assess the EDB code, excluded beneficiaries who self-identify with multiple races, and did not examine any changes to outcomes reporting as a result of inaccurate race and ethnicity identification.²⁰ Assessing and comparing the validity of both the EDB and RTI race codes would help identify whether the source of misclassification error in the validation population is due to response or algorithmic bias and inform opportunities for CMS to improve race and ethnicity information via improved survey collection or improved imputation methods. This is especially important for the Medicare Advantage (MA) program, which has experienced greater enrollment among racial/ethnic minorities than White beneficiaries in the last decade.²¹

To address this gap, our study had three primary objectives. First, we compare the performance of the EDB and RTI race codes in a nationally representative sample of Medicare Advantage beneficiaries. Second, we determine if misclassification of race and ethnicity leads to bias in outcomes reporting for hospitalization and readmission rates. Third, we see how performance varies among Medicare Advantage beneficiaries who self-report having multiple racial and ethnic identities.

2 | METHODS

2.1 | Study design and data sources

Our primary source of data was the Medicare Master Beneficiary Summary File (MBSF) from 2015 to 2017 which includes a record for every Medicare beneficiary each year and includes the EDB and RTI race/ethnicity codes. We linked these data by unique beneficiary identification number to two sources of self-report race/ethnicity from January 1, 2015 to December 31, 2017: (1) The Medicare Advantage Health Outcomes Survey (HOS), and (2) the Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys.

The HOS is a survey instrument used to measure patient-reported functional status health outcomes, primarily physical and mental health functioning, and is administered annually to random samples of 1200 Medicare beneficiaries, each of which are drawn from a managed care organization with an MA contract.^{22,23} The MA CAHPS is a survey instrument used to measure patient-reported experiences with health care services in various settings and is administered annually to a sample of 600 Medicare beneficiaries drawn from each MA contract.²⁴ Both surveys include questions that allow individual respondents to self-report their race and ethnicity.

For validating the classification performance of the EDB and RTI race/ethnicity variables, we perform a cross-sectional study to compare beneficiary race and ethnicity as documented in administrative Medicare data to race and ethnicity as self-reported by Medicare beneficiaries in survey responses.

For assessing the impact of misclassified race/ethnicity data on outcomes reporting, we used the Medicare Provider Analysis and Review (MedPAR) files from 2015 to 2017 to identify hospital admissions and unplanned readmissions for MA beneficiaries. The MedPAR files contain records of hospital inpatient stays for an estimated 92% of MA beneficiaries and can also be linked at the individual level.²⁵ We calculate annualized hospitalization and readmission rates among MA beneficiaries when grouped by race/ethnicity derived from the RTI race code compared to self-report.

All the datasets in this study were de-identified and considered secondary research exempt from IRB review. All analyses were conducted using STATA, version 17.

2.2 | Study population

The study population included all Medicare beneficiaries at least 18 years of age who were enrolled in a MA contract and completed either a HOS or CAHPS survey between 2015 and 2017. Beneficiaries who responded to a HOS or CAHPS survey and did not answer the questions on self-reported race and ethnicity were also considered incomplete respondents and dropped from the study population. To assess any potential nonresponse bias, we compared observable characteristics between beneficiaries who completed a survey with those who did not complete a survey in the study population.

2.3 | Definition of race/ethnicity variables in Medicare

In the MBSF, which is used to ascertain beneficiary race/ethnicity for most claim-based studies of the Medicare program, the EDB and RTI race variables are coded as single variables in which Medicare beneficiaries are classified into one of these six categories: “White,” “Black,” “Hispanic,” “Asian/Pacific Islander,” “American Indian/Alaska Native” or “Other.” Missing values are coded separately with the label “Unknown.”^{26,27}

For both the EDB and RTI codes, Hispanic ethnicity is not coded separately and categories are mutually exclusive, which restricts Medicare beneficiaries from being classified into more than one race/ethnicity group. Based on this classification system, the term “White” should be interpreted as referring to non-Hispanic White only, “Black” refers to non-Hispanic Black or African American only, “Asian/Pacific Islander” refers to non-Hispanic (AANHPI) only, “American Indian/Alaska Native” refers to non-Hispanic American Indian or Alaska Native (AIAN) only, and “Other” refers to non-Hispanic Other race. The term “Hispanic” refers to Hispanic, Any Race.

2.4 | Definition of self-reported race and ethnicity

In the CAHPS and HOS surveys, Medicare beneficiaries report their race and ethnicity in response to two questions: first asking if the beneficiary is Hispanic or Latino, and then asking what the beneficiary's race is, allowing for multiple responses to be selected.

To align with the race/ethnicity categories used in Medicare data, we considered beneficiaries responding to CAHPS or HOS to be self-reported Hispanic if they selected any of the response options with an affirmative "Yes" in response to the first question regarding Hispanic ethnicity, regardless of their response to the second question regarding race. Beneficiaries were considered self-reported AANHPI if they chose an Asian or Pacific Islander related race in either survey (eFigure 1 in the Supplement).

2.5 | Validation of the EDB and RTI race codes

We take two approaches to measuring the performance of the EDB and RTI race codes used for beneficiary race/ethnicity in Medicare data. First, for each of the Medicare defined race/ethnicity categories, we measure how accurately the EDB and RTI race codes binarily classify Medicare beneficiaries into or out of that category, which allows us to assess classification performance for Medicare beneficiaries with multiracial identities. We calculate and report sensitivity, specificity, and positive and negative predictive values for each race/ethnicity category.

Second, because Medicare race/ethnicity coding uses mutually exclusive race/ethnicity categories, we filter our study sample to the subset of beneficiaries who self-report only one race and measure the level of agreement in how single-race Medicare beneficiaries are classified between the Medicare defined race/ethnicity categories as documented in the variables compared to self-report. We calculate and report sensitivity, specificity, and positive and negative predictive values for each race/ethnicity category.

2.6 | Reporting of hospitalization, 30-day, and 90-day readmission outcomes

For each year of the 2015–2017 MedPAR files, we identified all-cause index hospital admissions and flagged subsequent admissions as unplanned 30-day and 90-day readmissions based on date of service and the CCS or ICD-10 codes attached to the claims, applying the same criteria for identifying unplanned readmissions as those used by CMS to publicly report hospital-level readmission measures.²⁸ Additional details are available in the Appendix (Data S1). Claims were linked back to the beneficiary file and rates for the 3-year period were calculated for the study population stratified by race/ethnicity as identified by the EDB and RTI race codes compared to self-report.

3 | RESULTS

3.1 | Study sample

We identified 1,020,515 Medicare Advantage beneficiaries in the MBSF who completed either a HOS or CAHPS survey between 2015 and 2017 and provided a response for self-reported race and ethnicity. Characteristics related to age, gender, and reasons for Medicare enrollment were similar overall between beneficiaries who responded and those who did not respond to either survey (eTable 1). However, beneficiaries who did not respond to the survey were also more likely to be from the South Census region (39.3% vs. 30.9%), less likely to be from the West Census region (21.2% vs. 25.1%), and less likely to be dual-eligible for Medicare and Medicaid (17.1% vs. 25.1%).

Table 1 compares the distribution of the Medicare defined race/ethnicity categories for the study sample based on information recorded in the EDB race code, the RTI race code, and beneficiary self-report. For the same sample of MA enrolled beneficiaries, we observe variation in the distribution of reported race/ethnicity based on the source of race/ethnicity information. When we use the EDB race code, we find non-Hispanic White beneficiaries overrepresented in the study sample by 6.3 percentage points compared to self-report (79.6% to 73.3%). The RTI race codes slightly underreport both Hispanic beneficiaries (11.6% to 12.2%) and non-Hispanic Asian American or Native Hawaiian Pacific Islander (AANHPI) beneficiaries (3.0% to 3.7%). However, both the EDB and RTI race codes underreport non-Hispanic American Indian Alaska Native (AIAN) beneficiaries in the study sample (0.3% and 0.3% to 2.5%). These differences between sources are diminished when we limit the study sample to beneficiaries who report a single-race category only, but this is due to a greater proportion of AIAN beneficiaries who report a multiracial identity and are excluded from the sample (eTable 4).

3.2 | Classification performance of the EDB and RTI race codes

Table 2 shows the validation of the EDB and RTI race codes compared with beneficiary self-reported race/ethnicity and measures how accurately a beneficiary is binarily classified as either being in or out of that category for each of the Medicare defined race/ethnicity categories. Both the EDB and RTI race codes perform well when binarily classifying beneficiaries as non-Hispanic Black or not non-Hispanic Black, with high sensitivities (96.0%, 95.6%) and specificities (99.1%, 99.4%). A sensitivity of 96.0% and 99.1% tells us that among beneficiaries who self-report being non-Hispanic Black in our study population, 96.0% and 99.1% of them are being correctly classified as non-Hispanic Black by the EDB and RTI race codes, respectively. A specificity of 95.6% and 99.4% tells us that among beneficiaries who self-report being not non-Hispanic Black in our study population, 95.6% and 99.4% of them are being correctly classified as not non-Hispanic Black by the EDB and RTI race codes, respectively.

TABLE 1 Distribution of race and Hispanic ethnicity among survey respondents, by Medicare EDB race code and RTI race code versus beneficiary self-report.

	EDB race code (N = 1,020,515)	RTI race code (N = 1,020,515)	Self-reported ^a	
			Multiracial (N = 1,020,515)	Single race only (N = 990,271)
White	812,503 (79.6%)	740,812 (72.6%)	747,925 (73.3%)	726,423 (73.4%)
Black	113,106 (11.1%)	109,413 (10.7%)	108,858 (10.7%)	100,327 (10.1%)
Hispanic	36,623 (3.6%)	117,847 (11.6%)	124,374 (12.2%)	120,702 (12.2%)
AANHPI	23,050 (2.3%)	30,558 (3.0%)	37,896 (3.7%)	34,099 (3.4%)
AIAN	2,943 (0.3%)	2,894 (0.3%)	25,086 (2.5%)	3,683 (0.4%)
Other	19,134 (1.9%)	7,824 (0.8%)	5,037 (0.5%)	5,037 (0.5%)
Unknown ^b	13,156 (1.3%)	11,167 (1.1%)	—	—

Note: Data Sources: Data linked between Medicare Master Beneficiary Summary File (MBSF), Medicare Advantage Health Outcomes Survey (HOS), and the Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS) from 2015 to 2017.

Abbreviations: AANHPI, non-Hispanic Asian American or Native Hawaiian Pacific Islander; AIAN, Non-Hispanic American Indian Alaska Native; EDB, Enrollment Database; RTI, Research Triangle Institute; White, non-Hispanic White; Black, non-Hispanic Black; Hispanic, Hispanic any race; Other, Non-Hispanic Other race.

^aIn the HOS and CAHPS, survey respondents are allowed to self-report more than one race category, which is represented in the column for “Self-reported, Multiracial” with row numbers that will sum to be greater than the total number of respondents. To compare with the Medicare race and ethnicity codes, which do not allow for multiracial identity, we include the subset of survey respondents self-reporting only one race category in the column for “Single race only”.

^bIn the EDB and RTI race codes, any missing values are coded with the value of “Unknown”. In the CAHPS and HOS surveys, any surveys with a missing response to the questions for self-reported race and Hispanic ethnicity are considered incomplete and dropped from the validation analysis.

TABLE 2 Validation of the Medicare EDB and RTI race codes compared to beneficiary self-reported race and Hispanic ethnicity, among all survey respondents (N = 1,043,559).

	Classification by EDB race code ^a				Classification by RTI race code ^a			
	Sensitivity ^b	Specificity ^c	PPV ^d	NPV ^e	Sensitivity ^b	Specificity ^c	PPV ^d	NPV ^e
White	97.6%	69.6%	89.8%	91.3%	96.6%	93.3%	97.5%	90.9%
Black	96.0%	99.1%	92.4%	99.5%	95.6%	99.4%	95.1%	99.5%
Hispanic	28.3%	99.8%	96.0%	90.9%	85.9%	98.8%	90.7%	98.1%
AANHPI	56.1%	99.8%	92.2%	98.3%	72.1%	99.7%	89.4%	98.9%
AIAN	8.7%	99.9%	74.3%	97.7%	8.7%	99.9%	75.1%	97.7%
Other	5.1%	98.1%	1.4%	99.5%	3.0%	99.2%	1.9%	99.5%

Note: Data Sources: Data linked between Medicare Master Beneficiary Summary File (MBSF), Medicare Advantage Health Outcomes Survey (HOS), and the Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS) from 2015 to 2017.

Abbreviations: AANHPI, non-Hispanic Asian American or Native Hawaiian Pacific Islander; AIAN, non-Hispanic American Indian Alaska Native; EDB, Enrollment Database; NPV, negative predictive value; PPV, positive predictive value; RTI, Research Triangle Institute; N, number of beneficiaries; White, non-Hispanic White; Black, non-Hispanic Black; Hispanic, Hispanic any race; Other, non-Hispanic Other race.

^aWe take a broad interpretation of the Medicare race codes and validate the predictive performance of the EDB and RTI race codes for each race/ethnicity category by measuring how accurately a beneficiary in our study sample is binarily classified as either in or out of that category. Because the current Medicare race/ethnicity codes do not allow for multiracial identity, this allows us to measure predictive performance for beneficiaries that self-report more than one race category. See eTable 2 (Supplemental material) for the full confusion matrices used to generate these metrics and Table 3 for validation of the predictive performance of the EDB and RTI race codes among survey respondents self-reporting only one race category.

^bSensitivity is defined as true positives/(true positives + false negatives).

^cSpecificity is defined as true negatives/(true negatives + false positives).

^dPositive predictive value is defined as true positives/(true positives + false positives).

^eNegative predictive value is defined as true negatives/(true negatives + false negatives).

For the classification of beneficiaries as non-Hispanic White or not non-Hispanic White, the EDB and RTI race codes have high sensitivities (97.6%, 96.6%), but the EDB code has lower specificity (69.6%, 93.3%). For Hispanic beneficiaries, the EDB code has very low sensitivity compared to the RTI code (28.3%, 85.9%), and both codes have

high specificity (99.8%, 98.8%). When classifying beneficiaries as non-Hispanic AANHPI, both the EDB and RTI race codes have lower sensitivity (56.1%, 72.1%), but high specificity (99.8%, 99.7%). However, both the EDB and RTI race codes have extremely low sensitivity for the classification of beneficiaries as non-Hispanic AIAN (8.7%, 8.7%).

When we assess the positive predictive values (PPV) and negative predictive values (NPV) of the Medicare race/ethnicity variables, we find that both EDB and RTI race codes have generally high PPV and NPV for beneficiaries classified as non-Hispanic White, non-Hispanic Black, Hispanic any race, and non-Hispanic AANHPI. Both variables have similar NPV but lower PPV for beneficiaries who are non-Hispanic AIAN, indicating that 74.3% and 75.1% of beneficiaries classified as non-Hispanic AIAN by the EDB and RTI race codes, respectively, are correct when compared to self-report.

In Table 3, we restrict the study sample to those beneficiaries self-reporting only one race category and consider the Medicare defined race/ethnicity categories as mutually exclusive categories, the performance of both the EDB and RTI race codes improves for the classification of non-Hispanic AIAN beneficiaries, but still reflects poor sensitivity (38.7%, 38.4%).

3.3 | Comparison of hospital admission rates, 30-day, and 90-day hospital readmission rates

Table 4 shows the annualized hospital admission, 30-day, and 90-day readmission outcomes in the period between 2015 and 2017 among non-Hispanic beneficiaries in the study sample, when based on the EDB and RTI race code versus beneficiary self-report. For non-Hispanic White and non-Hispanic Black beneficiaries, hospital admission and readmission outcomes are similar whether using the EDB or RTI race codes compared to beneficiary self-report. However, for Medicare beneficiaries who self-identify as non-Hispanic AIAN, using the EDB or RTI race code results in underreporting of hospital

admission rates (31.5% and 31.6% vs. 34.6%) and 30-day readmission rates (12.3% and 12.4% vs. 14.2%).

In Table 5, we also compare reported hospital admission and readmission outcomes among Hispanic beneficiaries in the study sample, when using the EDB and RTI race code versus beneficiary self-report. When comparing Hispanic beneficiaries with any race in aggregate, there appears to be little variation in hospitalization or readmission outcomes between the EDB and RTI race codes compared to beneficiary self-report. However, when we stratify outcomes among Hispanic beneficiaries by self-reported race, we find differences in hospital admission and readmission rates by self-reported race that are being lost in the aggregation of racial categories into the Hispanic any race category by the EDB and RTI race codes, including the higher rates of hospital admission, 30-day readmission, and 90-day readmission among beneficiaries who are self-reported Hispanic AANHPI (27.2%, 14.9%, 20.3%) and Hispanic AIAN (30.0%, 14.4%, 20.5%).

4 | DISCUSSION

Our study has three primary findings. First, while the RTI race/ethnicity imputation algorithm improves upon the validity of Medicare EDB race/ethnicity codes for Hispanic any race and non-Hispanic AANHPI beneficiaries, there remain disparate differences in performance by self-reported race/ethnicity, and current Medicare race/ethnicity codes should not be considered valid for identifying beneficiaries who identify as non-Hispanic AIAN. Second, the disproportionate nature of the misclassification errors introduced by the imputed race/ethnicity codes may lead to biased reporting of hospitalization

TABLE 3 Validation of the Medicare EDB and RTI race codes compared to beneficiary self-reported race and Hispanic ethnicity, among survey respondents self-reporting one race category only (N = 1,013,423).

	Classification by EDB race code ^a				Classification by RTI race code ^a			
	Sensitivity ^b	Specificity ^c	PPV ^d	NPV ^e	Sensitivity ^b	Specificity ^c	PPV ^d	NPV ^e
White	98.0%	92.6%	89.8%	69.5%	97.0%	92.0%	97.6%	93.4%
Black	98.0%	99.8%	92.5%	99.1%	97.6%	99.7%	95.2%	99.4%
Hispanic	28.6%	91.0%	96.1%	99.8%	86.5%	98.1%	90.8%	98.8%
AANHPI	60.7%	98.6%	92.4%	99.8%	77.9%	99.2%	89.5%	99.7%
AIAN	38.7%	99.8%	66.5%	99.9%	38.4%	99.8%	67.4%	99.9%
Other	5.1%	99.5%	1.4%	98.2%	3.0%	99.5%	2.0%	99.3%

Note: Data Sources: Data linked between Medicare Master Beneficiary Summary File (MBSF), Medicare Advantage Health Outcomes Survey (HOS), and the Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS) from 2015 to 2017.

Abbreviations: AANHPI, non-Hispanic Asian American or Native Hawaiian Pacific Islander; AIAN, non-Hispanic American Indian Alaska Native; EDB, Enrollment Database; NPV, negative predictive value; PPV, positive predictive value; RTI, Research Triangle Institute; N, number of beneficiaries; White, non-Hispanic White; Black, non-Hispanic Black; Hispanic, Hispanic any race; Other, non-Hispanic Other race.

^aWe take a strict interpretation of the Medicare race codes and validate the predictive performance of the EDB and RTI race codes for each race/ethnicity category by measuring how accurately a beneficiary in our study sample is classified between the six multiple categories. See eTable 3 (Supplemental material) for the full confusion matrices used to generate these metrics.

^bSensitivity is defined as true positives/(true positives + false negatives).

^cSpecificity is defined as true negatives/(true negatives + false positives).

^dPositive predictive value is defined as true positives/(true positives + false positives).

^eNegative predictive value is defined as true negatives/(true negatives + false negatives).

TABLE 4 Measures of annualized hospital admission, 30-day, and 90-day readmission rates from 2015 to 2017, among non-Hispanic beneficiaries as identified from EDB and RTI race codes versus self-report.

	EDB race code	RTI race code	Self-reported
Race/ethnicity	Non-Hispanic White	Non-Hispanic White	Non-Hispanic White
N	739,485	670,282	676,098
Hospital admission rate	30.3%	30.8%	30.6%
30-day readmission rate	11.8%	11.8%	11.8%
90-day readmission rate	17.2%	17.1%	17.1%
Race/ethnicity	Non-Hispanic Black	Non-Hispanic Black	Non-Hispanic Black
N	108,559	104,957	104,342
Hospital admission rate	31.3%	31.6%	31.5%
30-day readmission rate	14.4%	14.4%	14.5%
90-day readmission rate	21.4%	21.5%	21.6%
Race/ethnicity	Non-Hispanic AANHPI	Non-Hispanic AANHPI	Non-Hispanic AANHPI
N	22,186	29,237	36,213
Hospital admission rate	20.6%	20.1%	20.7%
30-day readmission rate	11.7%	11.7%	11.5%
90-day readmission rate	16.7%	16.6%	17.0%
Race/ethnicity	Non-Hispanic AIAN	Non-Hispanic AIAN	Non-Hispanic AIAN
N	2733	2684	23,762
Hospital admission rate	31.5%	31.6%	34.6%
30-day readmission rate	12.3%	12.4%	14.2%
90-day readmission rate	19.4%	19.3%	21.0%

Note: Data Sources: Data linked between Medicare Master Beneficiary Summary File (MBSF), Medicare Advantage Health Outcomes Survey (HOS), the Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS), and the Medicare Provider Analysis and Review (MedPAR) from 2015 to 2017.

Abbreviations: AANHPI, Asian American or Native Hawaiian Pacific Islander; AIAN, American Indian Alaska Native; EDB, Enrollment Database; RTI, Research Triangle Institute.

TABLE 5 Measures of annualized hospital admission, 30-day, and 90-day readmission rates from 2015 to 2017, among Hispanic beneficiaries as identified from EDB and RTI race codes versus self-report.

	EDB code	RTI code	Self-reported	Self-reported	Self-reported	Self-reported	Self-reported	Self-reported
Race and ethnicity ^a	Hispanic any	Hispanic any	Hispanic any	Hispanic White	Hispanic Black	Hispanic AANHPI	Hispanic AIAN	Hispanic Other
N	35,760	114,278	120,618	83,104	7625	3741	3872	26,353
Hospital admission rate	27.7%	25.9%	26.2%	25.7%	25.0%	27.2%	30.0%	27.7%
30-day readmission rate	13.1%	12.3%	12.3%	12.0%	12.1%	14.9%	14.4%	12.9%
90-day readmission rate	19.4%	18.4%	18.3%	17.9%	17.9%	20.3%	20.5%	19.3%

Note: Data Sources: Data linked between Medicare Master Beneficiary Summary File (MBSF), Medicare Advantage Health Outcomes Survey (HOS), the Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS), and the Medicare Provider Analysis and Review (MedPAR) from 2015 to 2017.

Abbreviations: AANHPI, Asian American or Native Hawaiian Pacific Islander; AIAN, American Indian Alaska Native; EDB, Enrollment Database; RTI, Research Triangle Institute.

^aCurrent Medicare race/ethnicity codes aggregate beneficiaries with Hispanic ethnicity into a single-race/ethnicity category and therefore do not allow for the stratification of reported outcomes for beneficiaries with Hispanic ethnicity by race. In the right-hand columns, we include reported outcomes stratified by both self-reported race and Hispanic ethnicity but are unable to report similar outcomes stratified by EDB or RTI race code for comparison.

and readmission outcomes by race/ethnicity. Third, there are differences in reported outcomes between beneficiaries by both race and Hispanic ethnicity that are not being measured due to the aggregation of race and Hispanic ethnicity into a combined category by current Medicare race/ethnicity coding.

4.1 | Misclassification of race and ethnicity in Medicare

Our findings contribute to the literature in several ways. First, while prior studies have examined the performance of the RTI algorithm, several used older data and some used race/ethnicity information for a population largely in a post-acute care setting. Our study generally finds poorer performance than these prior studies when using a nationally representative community dwelling sample of Medicare beneficiaries. We also build on prior literature by comparing algorithm performance among those with multiple racial or ethnic identities and show how the Medicare program may currently be misclassifying outcome rates due to inaccurate race/ethnicity prediction.

Our findings that rates of race/ethnicity misclassification trend higher among beneficiaries identifying as Hispanic any race, non-Hispanic AANHPI, and non-Hispanic AIAN are consistent with results from previous validation studies.¹⁸⁻²⁰ However, in this study, we find overall worse performance and a greater disparity in misclassification rates by race/ethnicity. Since other validation analyses used beneficiary race and ethnicity collected through home health visits and nursing home assessments, differences in performance could be attributed to differences in how beneficiaries self-report race and ethnicity when assessed in clinical care settings versus a research survey setting. There may also be changes in the racial and ethnic makeup of the Medicare beneficiary population over time, as expansion of the Medicare Advantage program saw greater enrollment among racial/ethnic minorities.²¹

Previous validation studies of the EDB and RTI race codes also excluded from analysis any beneficiaries with self-reported multiracial identity, which may overlook the current diversity of racial and ethnic identity and mask the disparate effects of race/ethnicity misclassification.¹⁸⁻²⁰ The exclusion of multiracial beneficiaries from analysis would disproportionately exclude beneficiaries identifying as AIAN who most often reported multiple race categories (eTable 4).

Comparing the performance of the EDB and RTI race codes provides additional context for our findings. While we find that the RTI race code improves upon the sensitivity of the EDB code for identifying beneficiaries who are Hispanic any race or non-Hispanic AANHPI, we find no differences in performance between the EDB and RTI race codes for identifying beneficiaries who are non-Hispanic AIAN, regardless of whether they report multiple race categories (Table 2) or a single-race category (Table 3). This suggests that current imputation methods are not enough for identifying non-Hispanic AIAN beneficiaries, and improvements in survey collection may be needed for this population.

4.2 | Subsequent bias in outcome reporting by race and ethnicity

We find differences in reported hospitalization and readmission outcomes among beneficiaries with Hispanic ethnicity stratified by self-reported race. These differences cannot be measured using current Medicare race/ethnicity codes alone. Instead, the aggregation of beneficiaries with Hispanic ethnicity into Medicare defined Hispanic any race category would result in underestimated outcomes reported for beneficiaries self-reported Hispanic Black and Hispanic AIAN. Previous studies of racial identity among Hispanics in the United States provide evidence that Hispanic ethnicity, which refers to cultural values, norms, and behaviors, is distinct from race, which refers to categorization of individuals on the basis of perceived differences in skin color and appearance.²⁹ While imperfect, the distinction between race and Hispanic ethnicity is important for understanding how bias, racial discrimination, culture, socioeconomic status, access to care, environmental factors, and genetics relate to racial/ethnic differences in health outcomes.³⁰

Medicare's current practices to combine Hispanic ethnicity with race categories and to exclude consideration of beneficiary multiracial identity, in addition to prediction errors inherent in the imputation of missing race/ethnicity by the RTI race code, help explain why we find significant misclassification of race/ethnicity among beneficiaries in our study. While there has been discussion on how imputation methods can introduce biases as a result of misidentification due to underrepresentation, the performance of an imputation algorithm is also dependent on the race/ethnicity categories the algorithm is trained on.²⁷ When beneficiaries are allowed to self-report multiracial identities, we find examples of how the algorithm, by design, performs poorly at distinguishing which single-race category to classify the beneficiary (eTable 5, eTable 6).

4.3 | Implications for health services research and policy

Our study suggests several important implications for health services researchers using the administrative Medicare race/ethnicity variables in their research. Our findings agree with previous validation studies that the EDB and RTI race codes should be considered valid for identifying beneficiaries who self-report as non-Hispanic White or non-Hispanic Black, due to the high sensitivities and specificities of the variables for both race and ethnicity categories, although the EDB codes have lower specificity for non-Hispanic Whites indicating a higher chance of false positives. Our findings also agree with previous validation studies that the EDB codes should not be considered valid for identifying beneficiaries who self-report as Hispanic any race, non-Hispanic AANHPI, or non-Hispanic AIAN due to low sensitivity. The RTI race codes should be preferred for identifying beneficiaries who self-report Hispanic any race, non-Hispanic AANHPI, but researchers should keep in mind that lower sensitivities of 85.9% and 72.1% suggest that the variables may disproportionately undercount

these beneficiaries compared to non-Hispanic White or non-Hispanic Black beneficiaries in their sample.

Both the EDB and RTI race codes should not be used to identify beneficiaries who self-report as non-Hispanic AIAN, as both variables have low sensitivity and are likely to result in undercounting for these beneficiaries. This assessment remains true whether the sample includes beneficiaries who are multiracial or self-report a single-race category only. We also find that these misclassification errors result in undercounting of hospital admission and readmission outcomes for non-Hispanic AIAN, which may indicate that caution is needed when using the Medicare race/ethnicity variables to measure outcomes stratified by race and ethnicity.

Our study also suggests several important policy considerations. First, CMS should regularly audit the classification performance of their race/ethnicity imputation algorithms and implement processes to validate against self-reported race/ethnicity. While the performance of the RTI imputation algorithm may have been sufficient when first developed, the current process of using imputation combined with administrative EDB data may no longer be sufficient.^{14,31} Recent advances in race/ethnicity imputation algorithms, including the Medicare Bayesian Improved Surname Geocoding (MBISG), combine Medicare administrative records with additional data elements to improve classification performance, but a race/ethnicity code derived from this algorithm is not included in datasets available to researchers, and early versions of the algorithm also performed poorly for identifying AIAN and multiracial beneficiaries, which underscores the need for continuous monitoring, validation, and improvement over time.³²⁻³⁴

When we compare beneficiaries who are correctly or incorrectly identified by Medicare race/ethnicity codes, we find no differences in age, sex, or reasons for Medicare enrollment, but we do observe differences in geographic location and Medicare-Medicaid dual eligibility (eTable 7, eTable 8). For both single-race and multiracial beneficiaries, those who are misclassified are less likely to reside in the South Census Region and more likely to reside in the West Census Region. While the difference in the South may be due to greater survey nonresponse in our study population (Table 2), the difference in the West suggests a need for CMS to improve race/ethnicity for beneficiaries residing in this region. In addition, those who are misclassified are more likely to be dual-eligible for Medicare and Medicaid, which suggests a need for CMS to improve race/ethnicity data for the dual-eligible population, which include opportunities for CMS to supplement Medicare data with race/ethnicity from Medicaid.

Second, this study highlights the need for CMS to revise their race/ethnicity categories and to consider separately identifying Hispanic ethnicity and allowing multiracial identity.³⁵ Allowing for greater diversity in the representation of race and ethnicity categories allows for greater granularity and detail in the outcomes and quality of care reporting necessary for understanding systematic differences in health and health care differences among Medicare beneficiaries.³⁶

4.4 | Limitations

Our study has several limitations. First, our sample is conditional on an MA beneficiary response to the HOS or CAHPS survey and providing self-reported race and ethnicity information. Our results may under- or over-estimate misclassification if beneficiaries with a higher likelihood of misclassification are more or less likely to complete the survey. While we are unable to measure misclassification among those who do not complete the surveys and self-report race/ethnicity, we find that beneficiaries who do not complete the survey are more likely to reside in the South Census region (eTable 1), indicating that our results may be less generalizable in the South due to higher non-response rates and more research is needed to improve survey collection of race and ethnicity in this region.

Second, our sample is limited to MA beneficiaries; however, previous literature has found MA disproportionately enrolls racial and ethnic minorities and may be an important population to study race/ethnicity classification.²¹ Third, we are limited by the questions available in the HOS or the CAHPS surveys which still might not capture all aspects of an individual's self-racial or ethnic identity. For the purposes of validating the Medicare race/ethnicity codes, we were required to aggregate some identities (such as Asian Americans, Native Hawaiians, and other Pacific Islanders into a single AANHPI category) for comparison purposes. Because this is a validation study, we do not intend to attribute any specific meaning to this race and ethnicity classification system but instead aim to measure the validity of using EDB or RTI race variables in Medicare data should one choose to use these race and ethnicity categories for studying the Medicare population.

5 | CONCLUSIONS

In this study, we find that current imputed race/ethnicity coding in administrative Medicare data disproportionately misclassifies and can lead to biased health outcomes reporting for Medicare beneficiaries who self-identify as non-Hispanic AANHPI, non-Hispanic AIAN, or Hispanic any race. While the imputation of missing race and ethnicity for Medicare beneficiaries has its merits, these findings suggest that ongoing evaluation is needed to ensure that algorithmic approaches to imputing race/ethnicity do not introduce additional bias when measuring disparities and inequities in health outcomes.³⁷

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (1R21MD016147, PI: David J. Meyers). We acknowledge Jeffrey Hiris for their technical support and colleagues at the Brown University School of Public Health for their feedback on this work. We also thank the reviewers and editors at Health Services Research for their comments on earlier versions of this paper.

ORCID

Andrew W. Huang  <https://orcid.org/0000-0001-7552-7091>

David J. Meyers  <https://orcid.org/0000-0002-4081-1751>

REFERENCES

1. Seshamani M, Fowler E, Brooks-lasure C. *Building on the CMS Strategic Vision: Working Together for a Stronger Medicare*. Health Affairs Forefront; 2022. doi:10.1377/forefront.20220110.198444/full/ Accessed April 26, 2022.
2. Brooks-Lasure C, Fowler E, Seshamani M, Tsai D. *Innovation at the Centers for Medicare and Medicaid Services: A Vision for the Next 10 Years | Health Affairs*. Health Affairs Forefront; 2021. Accessed April 26, 2022. doi:10.1377/forefront.20210812.211558/full/
3. Hughes DL. *CMS Innovation Center Launches New Initiative to Advance Health Equity*. Health Affairs Forefront; 2022. Accessed April 26, 2022. doi:10.1377/forefront.20220302.855616
4. Nerenz DR, Hunt KA, Escarce JJ. Health care organizations' use of data on race/ethnicity to address disparities in health care. *Health Serv Res*. 2006;41(4 I):1444-1450. doi:10.1111/j.1475-6773.2006.00613.x
5. Lauderdale DS, Goldberg J. The expanded racial and ethnic codes in the Medicare data files: their completeness of coverage and accuracy. *Am J Public Health*. 1996;86:712-716.
6. Filice CE, Joynt KE. Examining race and ethnicity information in Medicare administrative data. *Med Care*. 2017;55(12):e170-e176. doi:10.1097/MLR.0000000000000608
7. Arday SL, Arday DR, Monroe S, Zhang J. HCFA's racial and ethnic data: current accuracy and recent improvements. *Health Care Financ Rev*. 2000;21(4):107. Accessed October 2, 2021 /pmc/articles/PMC4194641/.
8. Zaslavsky AM, Ayanian JZ, Zaboriski LB. The validity of race and ethnicity in enrollment data for Medicare beneficiaries. *Health Serv Res*. 2012;47(3pt2):1300-1321. doi:10.1111/j.1475-6773.2012.01411.x
9. Fiscella K, Fremont AM. Use of geocoding and surname analysis to estimate race and ethnicity. *Health Serv Res*. 2006;41(4 I):1482-1500. doi:10.1111/j.1475-6773.2006.00551.x
10. Eicheldinger CR, Bonito A. More accurate racial and ethnic codes for Medicare administrative data. *Health Care Financ Rev*. 2008;29(3):27-42. Accessed September 17, 2021. /pmc/articles/PMC4195038/.
11. Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Relationship between quality of care and racial disparities in Medicare health plans. *JAMA*. 2006;296(16):1998-2004. doi:10.1001/jama.296.16.1998
12. Johnston KJ, Hammond G, Meyers DJ, Joynt Maddox KE. Association of race and ethnicity and Medicare program type with ambulatory care access and quality measures. *JAMA J Am Med Assoc*. 2021; 326(7):628-636. doi:10.1001/jama.2021.10413
13. Centers for Medicare & Medicaid Services. Medicare COVID-19 hospitalization trends. 2022 Accessed August 3, 2022. <https://www.cms.gov/research-statistics-data-systems/medicare-covid-19-hospitalization-trends>
14. Lines LM, Humphrey JL, Barch DH. Imputing race and ethnicity. *Med Care*. 2022;Publish Ah(5):351-356. doi:10.1097/mlr.0000000000001717
15. Irineo Cabrerros B, Agniel D, Martino SC, Damberg CL, Elliott MN. *Predicting Race and Ethnicity to Ensure Equitable Algorithms for Health Care Decision Making*. 2022;8:1153-1159. doi:10.1377/HLTHAFF.2022.00095
16. Ulmer C, McFadden B, Nerenz DR. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. National Academies Press (US); 2009. doi:10.17226/12696
17. Agency for Healthcare Research and Quality. Defining categorization needs for race and ethnicity data. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. National Academies Press; 2009 Accessed October 2, 2021. <https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldata3.html>
18. Jarrin OF, Nyandege AN, Grafova IB, Dong X, Lin H. Validity of race and ethnicity codes in Medicare administrative data compared with gold-standard self-reported race collected during routine home health care visits. *Med Care*. 2020;58(1):E1-E8. doi:10.1097/MLR.0000000000001216
19. Office of Inspector General. Data Brief: Inaccuracies in Medicare's Race and Ethnicity Data Hinder the Ability To Assess Health Disparities (OEI-02-21-00100). 2022 Accessed August 3, 2022. <https://oig.hhs.gov/oei/reports/OEI-02-21-00100.asp>
20. Zuckerman RB, Tarazi WW, Samson LW, et al. Quality of race and ethnicity data in Medicare. *Healthcare*. 2022;10(4):100662. doi:10.1016/j.hjdsi.2022.100662
21. Meyers DJ, Mor V, Rahman M, Trivedi AN. Growth in Medicare advantage greatest among black and hispanic enrollees. *Health Aff*. 2021;40(6):945-950. doi:10.1377/hlthaff.2021.00118
22. Jones N, Jones SL, Miller NA. The Medicare health outcomes survey program: overview, context, and near-term prospects. *Health Qual Life Outcomes*. 2004;2(1):1-10. doi:10.1186/1477-7525-2-33/TABLES/4
23. CMS Health Services Advisory Group. Sample Medicare HOS 2020 Cohort 23 Baseline Report. 2021 Accessed April 26, 2022. www.cms.gov/Research-Statistics-Data-and-Systems/Research/HOS
24. Centers for Medicare & Medicaid Services. Medicare Advantage and Prescription Drug Plan CAHPS Survey: Quality Assurance Protocols & Technical Specifications V6.0. 2015 Accessed April 26, 2022. http://www.ma-pdpcahps.org/globalassets/ma-pdp/quality-assurance/mapdp_cahps_survey_qapts-v6_0.pdf
25. Huckfeldt PJ, Escarce JJ, Rabideau B, Karaca-Mandic P, Sood N. Less intense Postacute care, better outcomes for enrollees in Medicare advantage than those in fee-for-service. *Health Aff (Millwood)*. 2017; 36(1):91-100. doi:10.1377/hlthaff.2016.1027
26. Research Data Assistance Center (ResDAC). Race/ethnicity (from Medicare EDB). Accessed April 26, 2022. <https://resdac.org/cms-data/variables/raceethnicity-medicare-edb>
27. Research Data Assistance Center (ResDAC). Research Triangle Institute (RTI) Race Code. Accessed April 26, 2022. <https://resdac.org/cms-data/variables/research-triangle-institute-rti-race-code>
28. Centers for Medicare & Medicaid Services. Quality Measures Fact Sheet Hospital-Wide All-Cause Unplanned Readmission Measure (NQF #1789) National Quality Strategy Domain: Communication and Care Coordination. 2019 Accessed April 26, 2022. <http://www.qualityforum.org/QPS/1789>
29. Borrell LN. Racial identity among Hispanics: implications for health and well-being. *Am J Public Health*. 2005;95(3):379-381. doi:10.2105/AJPH.2004.058172
30. Borrell LN, Elhawary JR, Fuentes-Afflick E, et al. Race and genetic ancestry in medicine—a time for reckoning with racism. *N Engl J Med*. 2021;384(5):474-480. doi:10.1056/NEJMms2029562
31. Grafova IB, Jarrin OF. Beyond black and white: mapping misclassification of Medicare beneficiaries race and ethnicity. *Med Care Res Rev*. 2021;78(5):616-626. doi:10.1177/1077558720935733
32. Haas A, Elliott MN, Dembosky JW, et al. Imputation of race/ethnicity to enable measurement of HEDIS performance by race/ethnicity. *Health Serv Res*. 2019;54(1):13-23. doi:10.1111/1475-6773.13099
33. Haas A, Adams JL, Haviland AM, et al. The contribution of first-name information to the accuracy of racial-and-ethnic imputations varies by sex and race-and-ethnicity among Medicare beneficiaries. *Med Care*. 2022;60(8):556-562. doi:10.1097/MLR.0000000000001732
34. Martino SC, Elliott MN, Klein DJ, et al. Disparities in the quality of clinical care delivered to American Indian/Alaska native Medicare advantage enrollees. *Health Aff (Millwood)*. 2022;41(5):663-670. doi:10.1377/hlthaff.2021.01830
35. Executive Office of the President Office of Management and Budget Office of Information and Regulatory Affairs. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. *Fed Regist*. 1997;62(210):58782-58790. Accessed April 26, 2022. <http://www.whitehouse.gov/WH/>
36. Kader F, Doan LN, Lee M, Chin MK, Kwon Simona C, Yi SS. *Disaggregating Race/Ethnicity Data Categories: Criticisms, Dangers, And Opposing Viewpoints | Health Affairs*. Health Affairs Forefront; 2022. doi:10.1377/forefront.20220323.555023

37. Nead KT, Hinkston CL, Wehner MR. Cautions when using race and ethnicity in administrative claims data sets. *JAMA Health Forum.* 2022;3(7):e221812. doi:[10.1001/JAMAHEALTHFORUM.2022.1812](https://doi.org/10.1001/JAMAHEALTHFORUM.2022.1812)

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Huang AW, Meyers DJ. Assessing the validity of race and ethnicity coding in administrative Medicare data for reporting outcomes among Medicare advantage beneficiaries from 2015 to 2017. *Health Serv Res.* 2023;58(5):1045-1055. doi:[10.1111/1475-6773.14197](https://doi.org/10.1111/1475-6773.14197)