

# Availability of Race and Ethnicity Data for Medicaid Beneficiaries

Racial and ethnic disparities in health care and health outcomes have been well documented over many years (MACPAC 2021a). Such disparities are particularly relevant for Medicaid given that more than half of the program's 73 million beneficiaries identify as Black, Hispanic, Asian American, or another non-white race or ethnicity (MACPAC 2021b). Addressing disparities and promoting equity in coverage, access, experience, and outcomes among historically marginalized and underserved populations will depend in part on having complete and systemically collected data by race and ethnicity.

This issue brief begins by describing federal data standards for capturing information on race and ethnicity and federal requirements for monitoring disparities in Medicaid. It then describes the availability of such information in Medicaid administrative data, specifically the Transformed Medicaid Statistical Information System (T-MSIS), and in federal household surveys that capture health care use and experiences of Medicaid beneficiaries, among others.

Although still in need of improvement, administrative data provide more comprehensive race and ethnicity data compared to federal survey data. All state Medicaid programs report race and ethnicity data, although the quality of the data varies by state. Thirty states meet the minimum data quality standards necessary for conducting analyses using 2019 race and ethnicity data. Although federal surveys include questions about race and ethnicity, the sample sizes for many groups are often too small to provide reliable estimates.

## Federal Standards

In response to the need for consistent data to enforce civil rights laws, the Office of Management and Budget (OMB) first established minimum federal standards in 1977 (OMB 1977).<sup>1</sup> These standards were designed to promote uniformity and comparability of data on race and ethnicity across the federal government in terms of collection, recordkeeping, and reporting.<sup>2</sup> These standards were updated in 1997 in recognition of the increasing diversity of the United States (OMB 1997, 1995). The OMB standard is required in all federally sponsored data collections, including U.S. Department of Health and Human Services (HHS) data collection initiatives.



The 1997 standards note that separate questions should be used whenever feasible for reporting race and ethnicity. The standards provide five minimum categories for race:

- American Indian or Alaska Native,
- Asian,
- Black or African American,
- Native Hawaiian or other Pacific Islander, and
- White.

The 1997 standards also established two categories for data on ethnicity (Hispanic or Latino and Not Hispanic or Latino).

OMB also noted that these standards should be considered the minimum for data collection; collection of more detailed information is permissible as long as any additional categories can be aggregated into the minimum standards. For example, a survey could provide detailed categories of Hispanic origin (e.g., Mexican, Cuban, Puerto Rican) as long as the categories can be aggregated into the OMB-specified Hispanic group. The OMB standards also note that self-identification is the preferred means of obtaining information about race and ethnicity and that individuals should be able to select more than one racial category (OMB 1997).<sup>3</sup>

## Application of federal standards to Medicaid and CHIP

The Patient Protection and Affordable Care Act (ACA, P.L. 111-148 as amended) included several provisions aimed at reducing health disparities. Specific requirements under Section 4302(a) directed the HHS Secretary to implement standards for the collection of data on five demographic categories: race, ethnicity, sex, primary language, and disability status. These standards affected federally conducted and supported health care or public health programs and activities, including Medicaid and CHIP, as well as data collection by self-report in federal surveys (§ 4302(b) of ACA).

HHS implemented updated minimum race and ethnicity standards in October 2011, drawing on recommendations of a workgroup consisting of representatives of HHS, OMB, and the U.S. Census Bureau. They mirror the five minimum racial categories and two categories for ethnicity set out in the 1997 standards, with additional granularity encouraged where possible. Selection of more than one race is permitted (HHS 2014, 2011).<sup>4</sup>

Section 4302(b) also requires continuing evaluation of this data collection, including reports to Congress by HHS four years after enactment and four years thereafter. In the 2014 report, HHS described its activities to implement Section 4302(b). These included application of the standards to T-MSIS, design of a survey of Medicaid beneficiaries on their experience of care, and development of a single streamlined application for Medicaid and the federally facilitated health exchanges (HHS 2014).



## Race and Ethnicity in Administrative Data

Administrative data on race and ethnicity are self-reported at the time of application and redetermination. Although most state Medicaid programs had long collected information about race and ethnicity as part of the enrollment process, the Centers for Medicare & Medicaid Services (CMS) incorporated many of the Section 4302 elements and questions when it developed the single streamlined application required under the ACA. The federal exchange and many state Medicaid programs began using this tool in October 2013 (HHS 2014).

States are required to report enrollee demographic characteristics, including race and ethnicity data, in T-MSIS whenever possible (CMS 2021a, CMS n.d.).<sup>5</sup> States follow national standards for reporting racial and ethnic groups in the Annual Demographic and Eligibility (DE) file of T-MSIS and use two separate elements to submit race and ethnicity data: the race code and the ethnicity code (CMS 2021e). As the DE file is large and complex, states must process the data from their own standardized format into a pre-approved CMS format before submission to a relational database (CMS 2021f). CMS then repackages the submitted data within the relational database into the research-ready T-MSIS analytic files (TAF).

There are 17 valid race code values and 7 valid ethnicity code values in the TAF. These two elements are combined to create an expanded race and ethnicity code with 20 valid values and a condensed race and ethnicity code with seven valid values (Table 1). Data are considered missing (i.e., null value) in the TAF if both data elements are missing or unknown, or if the race code is missing and the ethnicity code is missing or non-Hispanic.<sup>6</sup>

These data may be missing or unknown for several reasons. Individuals may choose not to disclose their race and ethnicity when applying for or renewing coverage. This may reflect concerns about discrimination or confusion regarding how the data might be used (Lukanen and Zylla 2021, 2020). It is important to note that respondents cannot be required to provide these data because they are not a factor for eligibility, and states have had varied success in encouraging people to fill out these fields. States may also not report complete data to CMS due to technical difficulties (CMS 2021c). CMS guidance to states includes improved T-MSIS race and ethnicity data as a priority area (CMS 2021b).



**TABLE 1.** Construction of the Condensed Race and Ethnicity Codes in the T-MSIS Analytic File

T-MSIS condensed race and ethnicity	T-MSIS race	T-MSIS ethnicity
White, non-Hispanic	White	Not of Hispanic or, Latino/a, or Spanish origin
Black, non-Hispanic	Black or African American	Not of Hispanic or, Latino/a, or Spanish origin
Asian, non-Hispanic	Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian, Asian Unknown	Not of Hispanic or, Latino/a, or Spanish origin
American Indian and Alaska Native (AIAN), non-Hispanic	American Indian or Alaskan Native	Not of Hispanic or, Latino/a, or Spanish origin
Hawaiian/Pacific Islander	Native Hawaiian, Guamanian or Chamorro, Samoan, other Pacific Islander, Native Hawaiian or other Pacific Islander/Unknown	Not of Hispanic or, Latino/a, or Spanish origin
Multiracial, non-Hispanic	Multiple races	Not of Hispanic or, Latino/a, or Spanish origin
Hispanic, all races	Any race(s)	Mexican, Mexican American, Chicano/a, Puerto Rican, Cuban, Another Hispanic, Latino/a, or Spanish origin, Hispanic or Latino/a Unknown

**Notes:** T-MSIS is the Transformed Medicaid Statistical Information System.

**Source:** CMS 2021c.

## Missing or incomplete data

MACPAC's analysis of 2019 T-MSIS data show high levels of missing or unknown data on race and ethnicity (Table A-1).<sup>7</sup> Specifically, only 19 states reported complete race and ethnicity data for at least 90 percent of their records and 3 states reported incomplete race and ethnicity data for at least 50 percent of their records. In addition, 15 states are missing both race and ethnicity values in over 10 percent of their records.<sup>8</sup>

A large number of states report missing race data. Twenty-three states are missing race data for between 10 and 30 percent of their enrollees and 9 states are missing race data for more than 30 percent. Only 19 states have more than 90 percent complete race data. Three states (Alabama, Massachusetts, and Utah) have over 50 percent incomplete race data.



Ethnicity data are more complete. Thirty-six states had more than 90 percent complete ethnicity data. Nine states had between 10 and 30 percent missing or unknown ethnicity data and 6 states had missing or unknown ethnicity data for more than 30 percent of their enrollees. No state had over 90 percent incomplete ethnicity data.

## Comparison with external benchmarks

Data that are reported on T-MSIS are not always consistent with other data sources, raising questions about their validity (Table 2). For example, in 13 states, reported enrollment among Hispanic individuals differed by more than 10 percentage points compared to those of the Medicaid population in the American Community Survey (ACS); in 4 states, this differed by more than 30 percentage points. Connecticut reported 0 percent of their Medicaid population as Hispanic despite ACS data indicating that the state's Hispanic Medicaid population is 33 percent. Enrollment among white, non-Hispanic individuals differed by more than 10 percent from the percentage reported in the ACS in 30 states. Similarly, enrollment among Black, non-Hispanic individuals differed by more than 10 percent compared to those of the ACS in 6 states (Table A-2).

**TABLE 2.** Comparison of 2019 T-MSIS Ethnicity Data with American Community Survey Data for Selected States

State	2019 T-MSIS data			Percentage Hispanic from ACS data	Percentage point difference between T-MSIS and ACS
	Percentage non-Hispanic	Percentage Hispanic	Unknown or missing		
Arizona	65.7%	0.8%	33.5%	44.8%	-44.0%
Colorado	51.1%	6.7%	42.2%	34.2%	-27.4%
Iowa	20.7%	50.6%	28.7%	8.9%	41.7%
Tennessee	40.3%	43.6%	16.1%	7.9%	35.7%
Connecticut	62.5%	0.0%	37.5%	33.0%	-33.0%

**Notes:** ACS is American Community Survey. Unknown indicates that the race or ethnicity variable was populated with a value corresponding to unspecified race or ethnicity, or an invalid value.

**Source:** MACPAC, 2022, analysis of 2019 T-MSIS and 2019 ACS data.

## Assessing data quality

MACPAC conducted data quality assessments for each state's race and ethnicity data using methods from the TAF DQ Atlas (CMS 2021c). There were two primary criteria for this assessment: the percentage of records with missing values and the number of combined race and ethnicity categories where 2019 TAF data differed from 2019 ACS values by more than 10 percent. We combined these two criteria to determine the usability of each state's race and ethnicity data, characterizing the data as low concern, medium concern, high concern, or unusable (Table 3).

Our assessment determined that 9 states had data of low concern, 21 states had data of medium concern, 17 states had data of high concern, and 4 states had unusable data. In total, 30 states (those with low and



medium data quality concerns) meet the minimum data quality standards necessary for conducting analyses with race and ethnicity data (Table A-2).

**TABLE 3.** Race and Ethnicity Code Data Quality Assessment Criteria

Percentage of records with missing values	Number of race and ethnicity categories where TAF differs from ACS by more than 10 percent	Data quality assessment
$x \leq 10$ percent	0	Low concern
$x \leq 10$ percent	1 or 2	Medium concern
$x \leq 10$ percent	3 or more	High concern
10 percent $< x \leq 20$ percent	0 or 1	Medium concern
10 percent $< x \leq 20$ percent	2 or more	High concern
20 percent $< x \leq 50$ percent	Any value	High concern
$x > 50$ percent	Any value	Unusable

**Notes:** ACS is American Community Survey. TAF is T-MSIS Analytic File.

**Source:** CMS 2021a.

## Race and Ethnicity in Survey Data

Multiple federal household surveys can be used to assess use of services and care experiences for Medicaid and CHIP beneficiaries. Among those frequently used by MACPAC are the National Health Interview Survey (NHIS), the Medical Expenditure Panel Survey, National Survey of Children’s Health, and the National Survey of Drug Use and Health. While these surveys can be useful in examining different aspects of health care use and experiences, none supports comprehensive analyses of coverage and access by race and ethnicity (Johnson et al. 2010).

Each of the federal surveys used by MACPAC measures race and ethnicity somewhat differently, although all follow the minimum standards established by OMB discussed above.<sup>9</sup> For example, while some surveys allow respondents to specify detailed categories of Hispanic origin (e.g., Mexican, Cuban, Puerto Rican, etc.), none allowed an ethnicity response other than Hispanic (e.g., Middle Eastern or North African) (SHADAC 2021).

The usefulness of federal survey data in providing estimates of access and service use by race and ethnicity is limited by the sample sizes (Table A-3). As a result, not all data that are collected are available in public use files, as they are often restricted for confidentiality reasons, and information on smaller populations is therefore not readily accessible. For example, for tables routinely included in MACStats using data from the NHIS, sample sizes are sufficient only to look at groups defined as Hispanic; white, non-Hispanic; Black, non-Hispanic; Asian, non-Hispanic; and other non-white, non-Hispanic (MACPAC 2021).

Furthermore, few federal surveys allow for state-level analyses, although some may support these analyses if multiple years of data are pooled (Bernard and Au-Yeung 2021). Lack of state-level estimates



are particularly problematic for examining Medicaid and CHIP given differences in state policies and in the underlying racial and ethnic makeup of state populations.

While the sampling strategies used in these surveys have allowed for examinations through oversampling, this is no longer done. For example, prior versions of the NHIS oversampled first Black, and then Hispanic and Asian respondents, creating the opportunity to calculate more robust estimates of what would otherwise be small populations. However, the current design does not oversample any racial or ethnic group at the household level.<sup>10</sup>

## Endnotes

<sup>1</sup> The OMB Directive No. 15 also implemented P.L. 94-311, which called for the analysis and publication of statistics on persons of Spanish origin or descent.

<sup>2</sup> Directive No. 15 provided a minimum set of standard categories and definitions for four racial categories (white, Black, American Indian or Alaskan Native, and Asian or Pacific Islander) and an ethnic category to indicate Hispanic origin. Collection of additional detail was permitted as long as the more detailed categories could be aggregated into minimum standards. The directive noted that it was preferable to collect data on race and ethnicity separately. Prior to its issuance in 1977, federal data collections used an assortment of definitions (OMB 1995, OMB 1977).

<sup>3</sup> The new standards were established for use in the 2000 decennial census. Other federal programs were expected to adopt them no later than January 1, 2003 for use in household surveys, administrative forms and records, and other data collections (OMB 1997). OMB issued a notice in the fall of 2016 requesting comments on revisions to the 1997 standards (OMB 2016). This notice appears not to have been finalized, most likely due to the change in administrations after the 2016 election. The Census Bureau conducted content testing on how to improve race and ethnicity information collection that indicated that a combined race and ethnicity question as well as a distinct Middle Eastern or North African category was the optimal design (Census Bureau 2021, Matthews et al. 2017). However, the Census Bureau announced that it would continue with the same dual question format as in the 2010 Census (Census Bureau 2018).

<sup>4</sup> As a result of a 1997 HHS data inclusion policy, the basic OMB standard on collecting race and ethnicity was already included in most HHS data collection initiatives (HHS 2011).

<sup>5</sup> The T-MSIS race and ethnicity data elements align with the 2011 HHS data collection standards. However, most states' Medicaid applications collect race and ethnicity data using the less-granular OMB standards (CMS 2021a).

<sup>6</sup> When both the race and ethnicity codes in the source T-MSIS data are missing, the race and ethnicity flag in TAF will be set to null. Additionally, if the ethnicity code is equal to zero (a valid value indicating non-Hispanic ethnicity) and the race code is missing in the source T-MSIS data, the race and ethnicity flag in TAF will be set to null. However, if the ethnicity code is missing and the race code is non-missing in T-MSIS, then the race and ethnicity code in TAF is set equal to the reported race code in T-MSIS.

<sup>7</sup> For our analysis, we followed the methods from the DQ Atlas (CMS 2021c). Therefore, if the race code was non-missing and the ethnicity code was missing in T-MSIS, we assigned the reported race and ethnicity code as race, non-Hispanic.



<sup>8</sup> A CMS analysis of the 2019 TAF found similarly high rates of missing data for these variables. Only 17 states reported complete race and ethnicity data for at least 90 percent of their records. This is in contrast to age and gender information which was complete for 99 percent of records for all states (CMS 2021d).

<sup>9</sup> Although “other” is not an official OMB designation, the Census Bureau and OMB also recommend including a “some other race” option.

<sup>10</sup> In prior years, certain groups were oversampled at the household level (Blacks in 1985-2015, Hispanics in 1995-2015, and Asians in 2006-2015) (CDC 2015, IPUMS undated).

## References

Bernard, S., and C. Au-Yeung. 2021. *Collection and availability of data on race, ethnicity, and immigrant groups in federal surveys that measure health insurance coverage and access to care.* Minneapolis, MN: State Health Access Data Assistance Center (SHADAC). <https://www.shadac.org/publications/fed-survey-data-REI>.

Centers for Disease Control and Prevention (CDC). 2015. National Health Interview Survey: Race and Hispanic origin information: Historical context. [https://www.cdc.gov/nchs/nhis/rhoi/rhoi\\_history.htm](https://www.cdc.gov/nchs/nhis/rhoi/rhoi_history.htm).

Centers for Medicaid and CHIP Services and Mathematica (CMCS and Mathematica). 2019. Technical assistance webinar: collecting and using stratified data for quality improvement in Medicaid and CHIP. July 11, 2019. <https://www.medicaid.gov/medicaid/quality-of-care/downloads/performance-measurement/core-set-stratification-webinar.pdf>.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2021a. Transformed Medicaid Statistical Information System (T-MSIS) coding blog – CMS technical instructions: Reporting race and ethnicity in the T-MSIS eligible file. <https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/109701>.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2021b. Transformed Medicaid Statistical Information System (T-MSIS) coding blog – CMS technical instruction: Overview of data quality T-MSIS priority items. <https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/54044>.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2021c. *DQ Atlas background and methods: Race and ethnicity.* Baltimore, MD: CMS. [https://www.medicaid.gov/dq-atlas/downloads/background\\_and\\_methods/TAF\\_DQ\\_Race\\_Ethnicity.pdf](https://www.medicaid.gov/dq-atlas/downloads/background_and_methods/TAF_DQ_Race_Ethnicity.pdf).

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2021d. *Explore by Topic (Explore Single Topic View), Race and Ethnicity.* Baltimore, MD: CMS. [https://www.medicaid.gov/dq-atlas/downloads/snapshots/2019/Release\\_1/topics/TAF\\_DQ\\_Snapshot\\_2019\\_R1.pdf](https://www.medicaid.gov/dq-atlas/downloads/snapshots/2019/Release_1/topics/TAF_DQ_Snapshot_2019_R1.pdf).

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2015. *Nonresponse bias analysis: Findings from the 2014-2015 Nationwide Adult Medicaid Consumer Assessment of Healthcare Providers and Systems.* Baltimore, MD: CMS. <https://www.medicaid.gov/medicaid/quality-of-care/downloads/performance-measurement/nonresponse-bias-analysis-brief.pdf>.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. n.d. CMS MACBIS T-MSIS reporting reminder: beneficiary demographics. <https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/51086>.





Farmer, B. 2020. As COVID-19 vaccine trials move at warp speed, recruiting black volunteers takes time. *National Public Radio*, September 11. <https://www.npr.org/sections/health-shots/2020/09/11/911885577/as-covid-19-vaccine-trials-move-at-warp-speed-recruiting-black-volunteers-takes>.

Integrated Public Use Microdata Series (IPUMS). n.d. NHIS sample design. [https://nhis.ipums.org/nhis/userNotes\\_sampledesign.shtml](https://nhis.ipums.org/nhis/userNotes_sampledesign.shtml).

Johnson, P.J., L. Blewett, and M. Davern, 2010. Disparities in public use data availability for race, ethnic, and immigrant groups: National surveys for healthcare disparities research. *Medical Care* 2010;48: 1122–1127. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3173979/pdf/nihms322952.pdf>.

Lukanen, E., and E. Zylla. 2021. *Collecting race, ethnicity, and language (REL) data on Medicaid applications: 50-state review shows wide variation in how states gather this information*. Princeton, NJ: State Health & Value Strategies. <https://www.shvs.org/resource/collection-of-race-ethnicity-language-rel-data-in-medicaid-applications-a-50-state-review-of-the-current-landscape/>.

Lukanen, E., and E. Zylla. 2020. *Exploring strategies to fill gaps in Medicaid race, ethnicity, and language data*. Princeton, NJ: State Health & Value Strategies. <https://www.shvs.org/exploring-strategies-to-fill-gaps-in-medicaid-race-ethnicity-and-language-data/>.

Matthews, K., J. Phelan, N.A. Jones, et al. 2017. *2015 national content test race and ethnicity analysis report*. Suitland, MD: U.S. Census Bureau. <https://s3.documentcloud.org/documents/4316468/2015nct-Race-Ethnicity-Analysis.pdf>.

Medicaid and CHIP Payment and Access Commission (MACPAC). 2021a. *Racial and ethnic disparities in Medicaid: An annotated bibliography*. Washington, DC: MACPAC. <https://www.macpac.gov/wp-content/uploads/2021/04/Racial-and-Ethnic-Disparities-in-Medicaid-An-Annotated-Bibliography.pdf>.

Medicaid and CHIP Payment and Access Commission (MACPAC). 2021b. Exhibit 2 in *MACStats: Medicaid and CHIP data book*. Washington, DC: MACPAC. <https://www.macpac.gov/wp-content/uploads/2015/01/EXHIBIT-2.-Characteristics-of-Non-Institutionalized-Individuals-by-Age-and-Source-of-Health-Coverage-2019.pdf>.

Ng, J.H., F. Ye, L.M. Ward, et al. 2016. Data on race, ethnicity, and language largely incomplete for managed care plan members. *Health Affairs* 36, no.3: 548–552. <https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.2016.1044>.

Office of Management and Budget (OMB). 2016. Standards for maintaining, collecting, and presenting federal data on race and ethnicity. Review and possible limited revision. *Federal Register* 81, no. 190 (September 30): 67398-67401. <https://www.govinfo.gov/content/pkg/FR-2016-09-30/pdf/2016-23672.pdf>.

Office of Management and Budget (OMB). 1997. Revisions to the standards for the classification of federal data on race and ethnicity. Notice of decision. *Federal Register* 62, no. 210 (October 30): 58782-58790. <https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf>.

Office of Management and Budget (OMB). 1995. Standards for the classification of federal data on race and ethnicity: Interim notice of review and possible revision of OMB's Statistical Policy Directive No. 15, race and ethnic standards for federal statistics and administrative reporting: summary and analysis of public comments and brief discussion of research agenda. *Federal Register* 60, no. 166 (August 28): 44674–44693. <https://www.govinfo.gov/content/pkg/FR-1995-08-28/pdf/95-20787.pdf>.

Office of Management and Budget (OMB). 1977. Directive No. 15: Race and ethnic standards for federal statistics and administrative reporting. [https://transition.fcc.gov/Bureaus/OSEC/library/legislative\\_histories/1195.pdf](https://transition.fcc.gov/Bureaus/OSEC/library/legislative_histories/1195.pdf).

State Health Access Data Assistance Center (SHADAC), University of Minnesota. 2021. Race, ethnicity, and nativity sample analysis for MACPAC. Minneapolis, MN: SHADAC.



U.S. Census Bureau, U.S. Department of Commerce. 2020. State oversampling in the National Survey of Children's Health: feasibility, cost, and alternative approaches. Suitland, MD: Census. [https://www.census.gov/content/dam/Census/programs-surveys/nsch/NSCH\\_State\\_Oversample\\_Summary\\_Document.pdf](https://www.census.gov/content/dam/Census/programs-surveys/nsch/NSCH_State_Oversample_Summary_Document.pdf).

U.S. Census Bureau, U.S. Department of Commerce. 2018. Census Bureau statement on 2020 Census race and ethnicity questions. January 26, 2018, press release. Suitland, MD: Census. <https://www.census.gov/newsroom/press-releases/2018/2020-race-questions.html>.

U.S. Department of Health and Human Services (HHS). 2014. *Report to Congress: Improving the identification of healthcare disparities in Medicaid and CHIP*. Washington, DC: HHS. <https://www.medicaid.gov/medicaid/quality-of-care/downloads/4302b-rtc-2014.pdf>.

U.S. Department of Health and Human Services (HHS). 2011. *Implementation guidance on data collection standards for race, ethnicity, sex, primary language, and disability status*. Washington, DC: HHS. <https://aspe.hhs.gov/system/files/pdf/76331/index.pdf>.



## APPENDIX: Availability of Race and Ethnicity Data for Medicaid Beneficiaries

**TABLE A-1.** Prevalence of Missing or Unknown Race and Ethnicity Values, FY 2019

State	Total (unknown race or ethnicity)	Unknown race	Unknown ethnicity	Unknown race and ethnicity
<b>Total</b>	<b>17.4%</b>	<b>17.4%</b>	<b>7.0%</b>	<b>7.0%</b>
Alabama	57.2%	57.2%	0.0%	0.0%
Alaska	6.8%	6.8%	0.0%	0.0%
Arizona	33.5%	33.5%	33.5%	33.5%
Arkansas	27.7%	27.7%	27.7%	27.7%
California	14.2%	14.2%	0.0%	0.0%
Colorado	42.2%	42.2%	7.1%	7.1%
Connecticut	37.5%	37.5%	11.7%	11.7%
Delaware	0.0%	0.0%	0.0%	0.0%
District of Columbia	27.1%	27.1%	23.8%	23.8%
Florida	14.8%	14.8%	0.0%	0.0%
Georgia	13.3%	13.3%	4.6%	4.6%
Hawaii	24.3%	24.3%	23.7%	23.7%
Idaho	0.1%	0.1%	0.1%	0.1%
Illinois	5.3%	5.3%	3.5%	3.5%
Indiana	9.9%	9.9%	9.9%	9.9%
Iowa	28.7%	28.7%	28.7%	28.7%
Kansas	9.4%	9.4%	9.4%	9.4%
Kentucky	16.3%	16.3%	16.3%	16.3%
Louisiana	35.4%	35.4%	35.3%	35.3%
Maine	9.1%	9.1%	0.1%	0.1%
Maryland	21.4%	21.4%	0.2%	0.2%
Massachusetts	51.0%	51.0%	49.8%	49.8%
Michigan	7.2%	7.2%	6.4%	6.4%
Minnesota	25.1%	25.1%	3.0%	3.0%
Mississippi	13.1%	13.1%	0.7%	0.7%
Missouri	6.6%	6.6%	3.2%	3.2%
Montana	12.6%	12.6%	0.0%	0.0%
Nebraska	11.1%	11.1%	6.4%	6.4%
Nevada	3.3%	3.3%	0.1%	0.1%
New Hampshire	11.3%	11.3%	0.0%	0.0%
New Jersey	11.2%	11.2%	0.0%	0.0%

State	Total (unknown race or ethnicity)	Unknown race	Unknown ethnicity	Unknown race and ethnicity
<b>Total</b>	<b>17.4%</b>	<b>17.4%</b>	<b>7.0%</b>	<b>7.0%</b>
New Mexico	1.6%	1.6%	0.7%	0.7%
New York	29.5%	29.5%	0.0%	0.0%
North Carolina	2.0%	2.0%	1.0%	1.0%
North Dakota	0.7%	0.7%	0.2%	0.2%
Ohio	7.6%	7.6%	0.0%	0.0%
Oklahoma	4.3%	4.3%	0.0%	0.0%
Oregon	15.6%	15.6%	0.3%	0.3%
Pennsylvania	5.8%	5.8%	0.6%	0.6%
Rhode Island	0.6%	0.6%	0.5%	0.5%
South Carolina	37.0%	37.0%	37.0%	37.0%
South Dakota	0.0%	0.0%	0.0%	0.0%
Tennessee	16.1%	16.1%	16.1%	16.1%
Texas	15.8%	15.8%	10.1%	10.1%
Utah	52.9%	52.9%	52.9%	52.9%
Vermont	17.2%	17.2%	17.2%	17.2%
Virginia	3.4%	3.4%	3.4%	3.4%
Washington	10.1%	10.1%	4.7%	4.7%
West Virginia	33.3%	33.3%	32.9%	32.9%
Wisconsin	16.7%	16.7%	0.1%	0.1%
Wyoming	21.7%	21.7%	0.0%	0.0%

**Notes:** Unknown indicates that the race or ethnicity variable was populated with a value corresponding to unspecified race or ethnicity, or an invalid value.

For our analysis, we followed the methods from the DQ Atlas (CMS 2021c). Therefore, if the race code was non-missing and the ethnicity code was missing in the Transformed Medicaid Statistical Information System (T-MSIS), we assigned the reported race and ethnicity code as race, non-Hispanic.

**Source:** MACPAC, 2022, analysis of 2019 T-MSIS data.

**TABLE A-2.** Data Quality Assessment of State Race/Ethnicity Codes, FY 2019

State	Missing (T-MSIS)	Difference in percentages (T-MSIS - ACS)						DQ assessment		
	Total	Hispanic	White, non-Hispanic	Black, non-Hispanic	Asian, non-Hispanic	AIAN, non-Hispanic	Hawaiian/Pacific Islander, non-Hispanic	Percent missing	Number of race and ethnicity categories (difference >  10% )	Assessment
Alabama	<b>57.2%</b>	-3.9%	-31.4%	-18.3%	-0.1%	-0.1%	0.1%	$x > 50\%$	2	Unusable
Alaska	<b>6.8%</b>	-1.5%	-0.7%	2.5%	3.5%	-7.6%	3.0%	$x \leq 10\%$	0	Low concern
Arizona	<b>33.5%</b>	-44.0%	8.0%	2.1%	0.2%	3.4%	-0.1%	$20\% < x \leq 50\%$	1	High concern
Arkansas	<b>27.7%</b>	11.7%	-17.0%	-19.0%	-0.6%	0.4%	0.3%	$20\% < x \leq 50\%$	3	High concern
California	<b>14.2%</b>	-8.2%	-2.4%	1.1%	-1.7%	0.0%	0.1%	$10\% < x \leq 20\%$	0	Medium concern
Colorado	<b>42.2%</b>	-27.4%	-13.0%	1.4%	0.1%	0.1%	0.1%	$20\% < x \leq 50\%$	2	High concern
Connecticut	<b>37.5%</b>	-33.0%	-1.0%	0.5%	-0.7%	0.3%	0.1%	$20\% < x \leq 50\%$	1	High concern
Delaware	<b>0.0%</b>	-1.7%	-4.9%	13.2%	-1.4%	-0.3%	0.5%	$x \leq 10\%$	1	Medium concern
District of Columbia	<b>27.1%</b>	-12.8%	-4.6%	-5.0%	-0.7%	-0.6%	0.0%	$20\% < x \leq 50\%$	1	High concern
Florida	<b>14.8%</b>	0.0%	-12.4%	2.1%	-0.6%	0.0%	0.0%	$10\% < x \leq 20\%$	1	Medium concern
Georgia	<b>13.3%</b>	-11.8%	-2.7%	5.7%	-0.4%	0.0%	0.1%	$10\% < x \leq 20\%$	1	Medium concern
Hawaii	<b>24.3%</b>	-8.3%	3.0%	0.9%	6.6%	1.1%	-1.7%	$20\% < x \leq 50\%$	0	High concern
Idaho	<b>0.1%</b>	-18.4%	22.1%	-0.9%	-0.6%	1.1%	-0.2%	$x \leq 10\%$	2	Medium concern
Illinois	<b>5.3%</b>	-1.7%	-9.8%	8.9%	0.0%	0.2%	0.1%	$x \leq 10\%$	0	Low concern
Indiana	<b>9.9%</b>	-0.8%	-11.8%	6.1%	-1.3%	-0.1%	2.3%	$x \leq 10\%$	1	Medium concern
Iowa	<b>28.7%</b>	41.7%	-78.1%	2.0%	9.2%	0.9%	-0.1%	$20\% < x \leq 50\%$	2	High concern
Kansas	<b>9.4%</b>	-0.5%	-9.0%	6.4%	-1.7%	1.5%	0.1%	$x \leq 10\%$	0	Low concern
Kentucky	<b>16.3%</b>	-1.3%	-12.2%	1.5%	-0.5%	-0.1%	0.0%	$10\% < x \leq 20\%$	1	Medium concern
Louisiana	<b>35.4%</b>	0.7%	-19.6%	-11.2%	-1.5%	-0.6%	0.0%	$20\% < x \leq 50\%$	2	High concern

State	Missing (T-MSIS)	Difference in percentages (T-MSIS - ACS)						DQ assessment		
	Total	Hispanic	White, non-Hispanic	Black, non-Hispanic	Asian, non-Hispanic	AIAN, non-Hispanic	Hawaiian/Pacific Islander, non-Hispanic	Percent missing	Number of race and ethnicity categories (difference >  10% )	Assessment
Maine	9.1%	-0.4%	-5.5%	1.3%	0.2%	-0.3%	0.0%	$x \leq 10\%$	0	Low concern
Maryland	21.4%	-2.7%	-13.4%	-0.7%	-0.1%	0.2%	0.1%	$20\% < x \leq 50\%$	1	High concern
Massachusetts	51.0%	-22.8%	-19.5%	-2.1%	-1.7%	0.0%	0.0%	$x > 50\%$	2	Unusable
Michigan	7.2%	-0.2%	-12.4%	10.3%	-0.9%	0.5%	0.1%	$x \leq 10\%$	2	Medium concern
Minnesota	25.1%	0.0%	-30.4%	6.3%	0.9%	2.0%	0.2%	$20\% < x \leq 50\%$	1	High concern
Mississippi	13.1%	-3.4%	-4.6%	-2.8%	-0.1%	0.2%	0.0%	$10\% < x \leq 20\%$	0	Medium concern
Missouri	6.6%	3.8%	-12.9%	7.1%	0.2%	-0.4%	-0.1%	$x \leq 10\%$	1	Medium concern
Montana	12.6%	-2.2%	-22.5%	0.5%	0.3%	13.8%	0.1%	$10\% < x \leq 20\%$	2	High concern
Nebraska	11.1%	0.3%	-13.1%	4.1%	1.2%	1.0%	-0.3%	$10\% < x \leq 20\%$	1	Medium concern
Nevada	3.3%	-6.4%	0.4%	7.3%	0.1%	0.3%	0.5%	$x \leq 10\%$	0	Low concern
New Hampshire	11.3%	-1.3%	-6.3%	0.0%	-0.1%	0.7%	0.0%	$10\% < x \leq 20\%$	0	Medium concern
New Jersey	11.2%	-6.4%	-3.4%	4.6%	-2.9%	0.2%	-0.1%	$10\% < x \leq 20\%$	0	Medium concern
New Mexico	1.6%	-8.1%	-0.8%	-0.2%	-0.4%	9.9%	0.0%	$x \leq 10\%$	0	Low concern
New York	29.5%	-11.1%	-15.1%	1.2%	-1.3%	0.3%	0.2%	$20\% < x \leq 50\%$	2	High concern
North Carolina	2.0%	-1.9%	-1.0%	7.0%	0.2%	-1.4%	0.1%	$x \leq 10\%$	0	Low concern
North Dakota	0.7%	-4.0%	-14.5%	6.4%	1.3%	13.2%	0.0%	$x \leq 10\%$	2	Medium concern
Ohio	7.6%	-2.0%	-9.4%	7.8%	-1.0%	0.2%	2.0%	$x \leq 10\%$	0	Low concern
Oklahoma	4.3%	-2.3%	-7.4%	4.5%	-0.1%	10.7%	0.3%	$x \leq 10\%$	1	Medium concern
Oregon	15.6%	16.9%	-27.6%	0.1%	-1.0%	1.0%	0.7%	$10\% < x \leq 20\%$	2	High concern
Pennsylvania	5.8%	1.4%	-13.7%	10.2%	0.4%	0.0%	0.0%	$x \leq 10\%$	2	Medium concern
Rhode Island	0.6%	-11.0%	-20.2%	-2.4%	-1.7%	-0.1%	41.6%	$x \leq 10\%$	3	High concern

State	Missing (T-MSIS)	Difference in percentages (T-MSIS - ACS)						DQ assessment		
	Total	Hispanic	White, non-Hispanic	Black, non-Hispanic	Asian, non-Hispanic	AIAN, non-Hispanic	Hawaiian/Pacific Islander, non-Hispanic	Percent missing	Number of race and ethnicity categories (difference >  10% )	Assessment
South Carolina	<b>37.0%</b>	-5.1%	-19.9%	-6.7%	-0.4%	0.0%	0.0%	20% < x ≤ 50%	1	High concern
South Dakota	<b>0.0%</b>	-3.0%	-17.5%	0.5%	0.3%	25.2%	0.2%	x ≤ 10%	2	Medium concern
Tennessee	<b>16.1%</b>	35.7%	-45.1%	-2.2%	-0.2%	-0.2%	-0.1%	10% < x ≤ 20%	2	High concern
Texas	<b>15.8%</b>	-5.9%	-8.0%	1.2%	-0.5%	0.1%	0.1%	10% < x ≤ 20%	0	Medium concern
Utah	<b>52.9%</b>	-5.1%	-44.0%	-0.3%	-1.2%	1.2%	-0.6%	x > 50%	1	Unusable
Vermont	<b>17.2%</b>	-1.8%	-13.3%	0.4%	0.5%	0.1%	0.0%	10% < x ≤ 20%	1	Medium concern
Virginia	<b>3.4%</b>	-7.4%	3.7%	5.5%	0.1%	0.3%	0.5%	x ≤ 10%	0	Low concern
Washington	<b>10.1%</b>	0.5%	-10.4%	2.7%	0.1%	2.1%	1.7%	10% < x ≤ 20%	1	Medium concern
West Virginia	<b>33.3%</b>	-1.6%	-27.1%	-1.0%	-0.2%	-0.1%	0.0%	20% < x ≤ 50%	1	High concern
Wisconsin	<b>16.7%</b>	1.4%	-21.8%	5.9%	0.6%	0.7%	0.0%	10% < x ≤ 20%	1	Medium concern
Wyoming	<b>21.7%</b>	-7.6%	-14.9%	0.6%	0.4%	5.6%	0.2%	20% < x ≤ 50%	1	Unusable
<b>Total</b>	<b>17.4%</b>	-4.4%	-11.6%	2.1%	-0.5%	0.6%	0.4%			

**Notes:** T-MSIS is the Transformed Medicaid Statistical Information System. ACS is the American Community Survey. Missing indicates that the race or ethnicity variable was populated with a value corresponding to unspecified race or ethnicity, or an invalid value.

**Source:** MACPAC, 2022, analysis of 2019 T-MSIS data.

**TABLE A-3. Unweighted Sample Size for Racial and Ethnic Groups in Selected Federal Surveys**

Survey (year)	Total	Hispanic <sup>1</sup>	White, non-Hispanic	Black, non-Hispanic	Asian, non-Hispanic	American Indian/Alaska Native, non-Hispanic	Native Hawaiian or other Pacific Islander, non-Hispanic	Other, multiple races, non-Hispanic
ACS (2019)	<b>391,865</b>	102,557	187,230	64,481	17,763	2,979	934	15,921
CPS (2020)	<b>22,783</b>	7,325	9,364	3,800	1,045	451	112	686
MEPS (2018)	<b>5,898</b>	2,301	1,746	1,249	107	31	176	288
NHANES (2017-2018)	<b>2,811</b>	806	760	767	243	N/A	N/A	235
NHIS (2019)	<b>5,445</b>	1,623	2,192	1,104	235	66	N/A	225
NSCH (2019)	<b>5,858</b>	1,207	3,088	806	207	70	28	452
NSDUH (2019)	<b>11,336</b>	3,228	4,390	2,416	344	284	86	588
PRAMS (2014)	<b>9,954</b>	2,071	3,742	2,548	447	551	14	581
SIPP (2018)	<b>9,775</b>	3,303	3,749	1,772	470	120	31	330

**Notes:**

N/A is not available. ACS is American Community Survey. CPS is Current Population Survey. MEPS is Medical Expenditure Panel Survey. NHANES is National Health and Nutrition Examination Survey. NHIS is National Health Interview Survey. NSCH is National Survey of Children's Health. NSDUH is National Survey of Drug Use and Health. PRAMS is Pregnancy Risk Assessment Monitoring System. SIPP is Survey of Income and Program Participation.

<sup>1</sup> The following surveys include additional detail on Hispanic ethnicity beyond yes/no: ACS, CPS, MEPS, NAMCAHPS, NHANES, NHIS.

<sup>2</sup> The ACS, CPS, MEPS, NAMCAHPS, and PRAMS have variables that provide detailed subcategories of Asian.

<sup>3</sup> Native Hawaiian or other Pacific Islander data are not identified in the NHIS public use file, but are available in the restricted data file.

<sup>4</sup> Medicaid coverage was defined based on the most recent date of data collection (e.g., December).

<sup>5</sup> The NSCH includes three categories of children's coverage (public, private, uninsured); we code children in public coverage as having Medicaid, because Medicare is much less common among children.

<sup>6</sup> Current coverage at the time of the data collection (PRAMS data include several additional coverage estimates, e.g., coverage at time of delivery).

<sup>7</sup> PRAMS data through 2017 available by CDC proposal.

**Source:** SHADAC, 2022, analysis of federal survey data.