DQATLAS

Background and Methods Resource

TOPIC
Race and Ethnicity

TOPIC AREA
Beneficiary Information

Summary

The TAF eligibility files include information on select demographic characteristics of beneficiaries enrolled in Medicaid or CHIP. This analysis examines the completeness of race and ethnicity information in the TAF. This analysis also examines how well the TAF data on race and ethnicity align with an external benchmark, the U.S. Census Bureau's American Community Survey.

Background

The T-MSIS Analytic Files (TAF) are research-optimized data on beneficiaries enrolled in Medicaid and the Children's Health Insurance Program (CHIP). The Annual Demographic and Eligibility (DE) file contains information on beneficiary demographic characteristics, including race and ethnicity.

States use two separate data elements to submit information on a beneficiary's race and ethnicity for T-MSIS: the race code, which has 17 categories, and the ethnicity code, which has 7 (Table 1). When eligibility records are created in TAF, these two source data elements are combined into an expanded race/ethnicity code that takes on 20 valid values and a condensed race/ethnicity code with 7 valid values. The expanded race/ethnicity code retains the detail of the source T-MSIS race code for beneficiaries with non-Hispanic or unspecified ethnicity and classifies beneficiaries with Hispanic ethnicity of any race in a single valid value. The condensed race/ethnicity code collapses the 19 non-Hispanic race codes into 6 broader valid values and retains the single valid value for all beneficiaries with Hispanic ethnicity. Both the expanded and condensed race/ethnicity codes are set to null (missing) in TAF when the T-MSIS source variables are missing or unknown, or when the race code is missing and the ethnicity code is either missing or equal to zero (non-Hispanic).

Table 1. Construction of the race/ethnicity codes in TAF

Data element	T-MSIS race	T-MSIS ethnicity	TAF expanded race/ethnicity	TAF condensed race/ethnicity
Step	1	1	2	3
Valid values	001 = White 002 = Black or African American 003 = American Indian or Alaskan Native 004 = Asian Indian 005 = Chinese 006 = Filipino 007 = Japanese 008 = Korean 009 = Vietnamese 010 = Other Asian 011 = Asian Unknown 012 = Native Hawaiian 013 = Guamanian or Chamorro 014 = Samoan 015 = Other Pacific Islander 016 = Native Hawaiian or Other Pacific Islander Unknown	0 = Not of Hispanic, Latino/a, or Spanish origin 1 = Mexican, Mexican American, Chicano/a 2 = Puerto Rican 3 = Cuban 4 = Another Hispanic, Latino/a, or Spanish origin 5 = Hispanic or Latino/a Unknown 6 = Ethnicity Unspecified	1 = White, non-Hispanic 2 = Black, non-Hispanic 3 = American Indian or Alaska Native, non-Hispanic 4 = Asian Indian, non-Hispanic 5 = Chinese, non-Hispanic 6 = Filipino, non-Hispanic 7 = Japanese, non-Hispanic 8 = Korean, non-Hispanic 9 = Vietnamese, non-Hispanic 10 = Other Asian, non-Hispanic 11 = Asian Unknown, non-Hispanic 12 = Multi-Asian, non-Hispanic 13 = Native Hawaiian, non-Hispanic 14 = Guamanian or Chamorro, non-Hispanic 15 = Samoan, non-Hispanic 16 = Other Pacific Islander, non-Hispanic 17 = Native Hawaiian or Other Pacific Islander Unknown, non-Hispanic 18 = Multi-Islander, non-Hispanic 19 = Multiracial, non-Hispanic 20 = Hispanic, any race	1 = White, non-Hispanic 2 = Black, non-Hispanic 3 = Asian, non-Hispanic 4 = American Indian and Alaska Native, non-Hispanic 5 = Hawaiian/Pacific Islander 6 = Multiracial, non-Hispanic 7 = Hispanic, all races

The race and ethnicity codes in the source T-MSIS data are monthly variables. During the construction of the annual TAF DE file, the race/ethnicity code is converted to an annual variable using the "last-best" method. The last-best method selects a variable's value from the most recent month for which a non-missing valid value exists. If no valid value was available in any month, TAF uses the last-best value from a previous calendar year. Thus, the DE file contains one expanded and one condensed race/ethnicity value per beneficiary in a given year.

Although states are expected to report the information that they receive on both race and ethnicity in T-MSIS, some states may not submit complete information because the data were not collected or technical difficulties arose in reporting. States may not have complete data on race and ethnicity because most do not require beneficiaries to disclose this information on enrollment forms, following the guidance from the Office of Management and Budget that establishes self-identification as the preferred means of obtaining this information. [1] States also vary in how they ask beneficiaries to identify ethnicity on enrollment forms, which may result in incomplete or inconsistent information on ethnicity in TAF for a state. [2]

Missing race and ethnicity information may affect all eligibility records in the DE equally (that is, the missingness may be randomly distributed), or the missing information may be concentrated among certain racial and ethnic groups. As a result, examining the proportion of DE records with missing data is not sufficient for understanding how usable this data element is for analytic purposes. This data quality assessment measures both the level of missing data and the extent to which the distribution of non-missing data aligns with an external benchmark, the American Community Survey (ACS).

Methods

For this analysis, we examined the condensed race/ethnicity code (RACE_ETHNCTY_FLAG) on non-dummy enrollment records [3] in the TAF DE file. [4] We tabulated the proportion of records that fell into each of the five valid race/ethnicity categories, as well as the proportion with missing values. We did not look for invalid values because the TAF production process recodes all invalid values and invalid combinations of race and ethnicity to a null value in TAF. [5]

To construct the benchmark, we used the ACS 5-year estimates [6] Public Use Microdata Sample (PUMS) [7] for a given year. The ACS data, which are collected annually from a nationally representative random sample of households, contains information on self-reported race, ethnicity, and health insurance coverage. After pulling the ACS microdata from PUMS, we selected all individuals who reported having "Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability" at the time of the survey. For individuals in that group, we calculated the percentage who were in the five race/ethnicity categories (Table 2).

ACS data is used by many stakeholders, including by federal and state government agencies for policy and program funding activities, and are considered a highly reliable source of demographic data. However, self-reporting of health insurance coverage for the ACS often results in an undercount of the number of Medicaid beneficiaries who appear in administrative data. Therefore, in this data quality assessment, we compare the percentage of Medicaid beneficiaries in each race/ethnicity group in TAF to the comparable distribution in the ACS, rather than comparing the count of individuals in each category.

Table 2. Race and ethnicity categories in TAF and ACS

Race/ethnicity category	RACE_ETHNCTY_FLAG value in TAF [a]	Combination of race and Hispanic variables in the ACS
White, non-Hispanic	1 = White, non-Hispanic	White alone, non-Hispanic
Black, non-Hispanic	2 = Black, non-Hispanic	Black alone, non-Hispanic
Asian, non-Hispanic	3 = Asian, non-Hispanic	Asian alone, non-Hispanic
Hispanic, all races	7 = Hispanic, all races	Hispanic, all races
All other races, non-Hispanic	4 = American Indian and Alaska Native, non-Hispanic 5 = Hawaiian/Pacific Islander 6 = Multiracial, non-Hispanic	• American Indian alone • Alaska Native alone • American Indian and Alaska Native tribes specified; or American Indian or Alaska native, not specified and no other race • Native Hawaiian and Other Pacific Islander alone • Some other race alone • Two or more races

[a] For the race/ethnicity flag TAF, a "non-Hispanic" value may indicate that (1) the beneficiary is not Hispanic or (2) the beneficiary's ethnicity or Hispanic status is not reported.

We used two criteria to assess each state's race/ethnicity data: (1) the percentage of enrollment records with missing data and (2) how well the percentage of beneficiaries in each of the five race/ethnicity categories aligned with the ACS (Table 3).

Table 3. Criteria for DQ assessment of race/ethnicity code

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

1. The OMB Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity were last revised in 1997. The 1997 revisions can be found at https://www.whitehouse.gov/wp-content/uploads/2017/11/Revisions-to-the-Standards-for-the-Classification-of-Federal-Data-on-Race-and-Ethnicity-October30-1997.pdf . See the Census Bureau website for a concise list of OMB race categories: https://www.census.gov/topics/population/race/about.html .

2. For example, some states provide beneficiaries with options to identify as "Hispanic/Latinx" or "Not Hispanic/Latinx" on enrollment forms, while other states list options for ethnicity without an explicit option to self-identify as "Not Hispanic/Latinx". Therefore, it may be difficult in the latter scenario to determine whether the beneficiary is non-Hispanic or if their Hispanic status is not reported.

3. We excluded DE records that represent beneficiary IDs that are present on claims but were not included in the eligibility records submitted by the state (those with MISG_ELGBLTY_DATA_IND = 1).

4. This analysis used the TAF data that were released as TAF Research Identifiable Files (RIF). During the transformation into RIF, some TAF data elements were suppressed, changed, or renamed. Additional details are available on the DQ Atlas Resources page , and a crosswalk of variable names can be found in the guide "Production of the TAF Research Identifiable Files."

5. When both the race and ethnicity codes in the source T-MSIS data are missing, the race/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/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/ethnicity code in TAF is set equal to the reported race code in T-MSIS.

6. ACS 5-year estimates are more reliable and complete than ACS 1-year estimates and the Current Population Survey, as it includes smaller geographic areas and has a larger sample size. For more details, see: https://www.census.gov/programs-surveys/acs/guidance/estimates.html and https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677056/

7. https://data.census.gov/mdat/#/