DQATLAS

Background and Methods Resource

Medicaid.gov
Keeping America Healthy

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.

Related topics include:

- Age
- Gender
- Income
- ZIP Code

Background

The T-MSIS Analytic Files (TAF) are an enhanced set of data on beneficiaries in Medicaid and the Children's Health Insurance Program (CHIP). These data include select demographic characteristics, which are critical to understanding those whom the programs serve. This data quality assessment examines the completeness and face validity of selected demographic variables in the TAF annual Demographic and Eligibility (DE) file.

We focused on five key demographic variables in the DE file that are important for analytic purposes: age group, gender, income, race/ethnicity, and ZIP code. Table 1 lists these variables, along with their data element names in the DE file and a brief description.

Table 1. Key demographic variables in the DE file

Variable	TAF Data Element Name	Description
Age group	AGE_GRP_FLAG	A beneficiary's age in years during the last month of enrollment in the calendar year, or as of the date of death if within the calendar year, grouped into 10 age categories
Gender	GNDR_CD	A beneficiary's biological sex
Income	INCM_CD	A code indicating a family's income level category
Race/Ethnicity	RACE_ETHNCTY_FLAG	A code indicating a beneficiary's race and ethnicity
ZIP code	ELGBL_ZIP_CD	The ZIP code corresponding to a beneficiary's residence. When residence ZIP code is unavailable, the ZIP code corresponding to the mailing address is used instead

Note: The age group, gender, income, race/ethnicity, and ZIP code variables in the DE file were each constructed using a "last-best" method for selecting values from the monthly TAF Beneficiary Summary Files (BSF). The "last-best" method selects a variable's value from the most recent month in the BSF for which a non-missing value exists. If a value was missing in the BSF for all months in a given year, TAF uses the "last-best" value from a previous calendar year.

While states are expected to collect and report information on all of these data elements in T-MSIS, some states may not report complete information due to not collecting the data or due to technical difficulties in reporting. Some states may not have complete data on race and ethnicity because they follow the guidance from the Office of Management and Budget (OMB) that establishes self-identification as the preferred means of obtaining this information, and not all beneficiaries disclose this information. [1]

Methods

We used the DE file to calculate the percentage of non-dummy [2] enrollment records that have complete data, which we define as valid, non-missing values. [3] For the age group, gender, income, and race/ethnicity variables, all non-missing values represent valid values because in the creation of TAF, all invalid values for categorical variables are recoded to null. We therefore considered non-missing values to represent complete data for these variables. [4] We also examined the ZIP code of residence on the enrollment record. For ZIP code, we considered complete data to include all values other than missing values or any 0-filled, 8-filled, or 9-filled values, which also represent missing data. [5]

We assessed data quality for each of these data elements based on the percentage of enrollment records with missing data (Table 2).

Table 2. Criteria for DQ assessment of key demographic variables

Percentage of records with missing values	DQ assessment
x ≤ 10 percent	Low concern
10 percent < x ≤ 20 percent	Medium concern
20 percent < x ≤ 50 percent	High concern
x > 50 percent	Unusable

We also tabulated the distribution of valid values for age group and gender to allow users to check for face validity. Table 3 provides information about how the age group data element was combined into three age categories.

Table 3. Collapsing of age groups into categories

Age group value in TAF	Description	Assigned age category
1	Age <1	Age 0–18
2	Age 1–5	Age 0–18
3	Age 6–14	Age 0–18
4	Age 15–18	Age 0–18
5	Age 19–20	Age 19–64
6	Age 21–44	Age 19–64
7	Age 45–64	Age 19–64
8	Age 65–74	Age 65+
9	Age 75–84	Age 65+
10	Age 85–125	Age 65+

The distribution of age group and gender is presented by state and Medicaid expansion status of the state. In states that opted to expand Medicaid, the expansion population is relatively large, and the age and gender distribution systematically differs from the traditional Medicaid population. We did not stratify based on other optional coverage groups because the populations are not as large and do not have the same impact on the age and gender distribution for a state's Medicaid and CHIP population.

We did not analyze income data beyond missing rates because there are no available benchmarks by which to compare the findings in the TAF. Furthermore, because Medicaid policies regarding income eligibility vary by state, checking for consistency across states is not feasible. We also did not analyze race/ethnicity beyond missing rates, although it is a known issue that some states do not report certain race/ethnicity groups that comprise a substantial proportion of the state's population. TAF users interested in this data element may want to further assess its validity by comparing TAF distributions to external benchmarks such as the American Community Survey.

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 and ethnicity categories: https://www.census.gov/topics/population/race/about.html .

2. As part of the creation of the TAF RIF, dummy records are added to the DE file that represent beneficiary IDs that are present on claims but were not included in the eligibility records submitted by the state. These dummy records can be identified using the data element MISG_ELGBLTY_DATA_IND.

3. 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."

4. Age group is equal to null if the source value from T-MSIS for that variable is missing, unknown, or not on the valid value list or within the range of valid values; gender is equal to null if the source value is missing, unknown, not on the valid value list or within the range of valid values, or equal to U; income is equal to null if the source value is missing, unknown, not on the valid value list or within the range of valid values, or equal to 88 or 99; race/ethnicity is equal to null if the source value is missing, unknown, not on the valid value list or within the range of valid values, or equal to 8.

5. ZIP code is equal to null if the source value for residence ZIP code and the mailing address ZIP code are missing, unknown, or invalid. Typically 0-filled, 8-filled, and 9-filled values indicate missing or unknown values. As a check on this decision, we confirmed that there are no valid residential 00000, 88888, or 99999 ZIP codes in the United States Postal Service system.