

# Open Source Code for the T-MSIS Analytic Files (TAF)

**Introduction:** The concept of “open source” has emerged as a cornerstone of modern software development, fostering innovation, collaboration, and transparency. At its core, open source code represents a philosophy that values the collective contribution and wisdom of a diverse community of users. By choosing an open source approach for our project, we hope to support a model that accelerates problem-solving and innovation, inviting researchers from across the U.S. to refine and enhance the work.

In the context of TAF data, open source code may hold several benefits for researchers. Given many researchers have similar needs (e.g., generating a reliable variable for MCO enrollment), it reduces redundant work by providing solutions or resources that can be used by anyone. Open source code can also provide standardization – for example, creating a standard definition of a pregnant enrollee or enrollee with opioid use disorder – minimizing the differences in definitional approaches across studies and increasing the reliability of our science. Open source code also serves as an opportunity to offer solutions to common problems across TAF data, such as using longitudinal data to impute missing demographic information.

## Repository

The Medicaid Data Learning Network group has created a GitHub repository, which can be found [here](#). This repository’s purpose is to assist researchers who work with data from the TAF. The repository contains a number of functions that perform common research tasks using the TAF data, such as providing lists of beneficiaries with certain conditions, evaluating data quality, constructing measures, and so on.

Using this repository provides value to researchers by:

- Allowing those working with TAF to utilize easy-to-use functions to perform frequently needed data preparation and manipulation functions rather than writing time-consuming functions themselves.
- Providing consistent and standardized procedures and definitions.

## Setup & Requirements

As noted in the repository, to use the taf.functions package, researchers will first need to install R and the ‘devtools’ package for R (the devtools package is necessary to install packages that are on GitHub rather than Cran). We have provided step-by-step instructions on the GitHub repository.

## Reliance on R Language

Our current efforts have focused on code written in R. However, we recognize that not all researchers use R. We use R because it is free, meaning that anyone can download and install the system. Furthermore, most programming languages (including Stata) allow you to run code written in another language. Thus, it will be possible to use these R functions while still working in your preferred programming language. We provide step-by-step instructions on how to use Stata to call the taf.functions from R in the repository.

## List of Current Functions

The repository includes the following functions:

**get\_apr\_data\_quality\_measures()** This function replicates **DQ Atlas**’ analysis of data quality in the Annual Provider (APR) files.

The function takes data from the ‘taf\_prvdr\_base’ file as input, and returns data similar to what is made available in the DQ Atlas. However, the DQ Atlas does not publish quality assessments for every year/release combination for which data are available. Thus, this function can re-create this analysis for years/releases where the DQ Atlas data are unavailable or on incomplete subsets of taf\_prvdr\_base data that may interest researchers.

**get\_icd10\_codes\_for\_opioid\_use\_disorder()** This function returns all ICD10 codes for opioid use disorder as a dictionary. ICD10 codes were chosen and named based on the information [here](#).

**get\_taf\_opioid\_use\_disorder\_patients()** This function returns a dataframe with the `msis_id`, `bene_id`, and `state_cd` of every beneficiary in the input data with an opioid-related disorder, defined using the National Quality Forum (NQF 12) approach: “enrollees with OUD diagnoses (using International Classification of Diseases, Ninth Revision [ICD-codes 304.0x, 305.5x and ICD-10 codes F11.xxx) recorded during an inpatient stay or a visit to an outpatient facility or office.”

The specific implications of employing the NQF approach when working with TAF are that inpatient and outpatient diagnostic codes will be considered. This means that all columns labeled ‘`dgns_cd`’ in the `taf_inpatient_header` and `taf_other_services_header` as outpatient codes are stored in the ‘other services’ header file. However, long term care diagnostic codes will not be considered.

**fill\_in\_missing\_demographic\_data()** Many enrollees fill out demographic information for one year but not others. This function fills in missing demographic data in instances where data are missing in some years but available and consistent in others. If information is missing in all years, this function does nothing, and if information is missing in some years and information in the other years is not consistent, this function does nothing.

### Ongoing work

We are currently working on creating a file that would substantially increase the utility of the TAF Annual Managed Care Plan (APL) Files. The TAF APL files include human-entered plan names, which are inconsistent and error-prone. In addition, there are more plan IDs than plan names, so plan IDs do not solve plan name inconsistencies. There is also a lack of data about the plans and how they roll up into larger parent organizations (United, Centene, etc). The analysts at Oregon Health & Science University are currently working on a cross-walk that will match TAF managed care plan names to standard plan names used by data.medicaid.gov, which can then be used to match plans to parent companies.

The Medicaid Data Learning Network (MDLN) offers researchers using T-MSIS Analytic Files (TAF) the opportunity to foster peer-shared learning and work collaboratively to develop key methodological standards to ensure that Medicaid research using the TAF data is high quality, timely, and impactful. With support from The Commonwealth Fund and the Robert Wood Johnson Foundation, the MDLN resides within AcademyHealth’s Evidence-Informed State Health Policy Institute (ESHPI) and is guided by a project team including AcademyHealth ESHPI staff and three TAF research faculty, Drs. Sarah Gordon, John McConnell, and William Schpero. The MDLN’s ultimate goal is to improve the quality of the TAF data over time, expand opportunities for health services researchers to use Medicaid claims data, and increase the number of researchers engaged in Medicaid-focused research.

You can learn more about the MDLN and access publications on AcademyHealth’s [website](#).