Managed Care Crosswalk Codebook

1. Introduction

The purpose of the managed care crosswalk is to facilitate the use and analysis of managed care data in the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) data.

Data on managed care organizations (MCOs) and their plans are reported by three sources known to CHSE:

1. **CMS "Managed Care Enrollment by Program and Plan"** **datasets**
2. **KFF "Total Medicaid MCO Enrollment"** **datasets**
   Data for years prior to 2022 must be accessed via the Internet Archive, a digital archive of web sites as they historically appeared.
3. **TAF Data**: Available via ResDAC, requiring a formal request.

While KFF uses CMS data, it also performs independent analyses that cannot be fully replicated using CMS data alone, effectively making it a separate source.

These sources do not share standardized names or ID numbers for managed care entities, making data integration challenging. This crosswalk provides a structured approach to merging data from these sources, enhancing the ability of researchers to analyze managed care enrollment and organization-level details across datasets.

The **purpose of this codebook** is to:

- Explain the meaning and source of each column in the crosswalk.
- Guide researchers in using the crosswalk effectively.
- Describe the methodology used to create the crosswalk.

2. Crosswalk Variables

This section defines all the variables in the crosswalk. Each row in the crosswalk represents an unique managed care plan, based on a combination of state and plan name variables (further described in Section 3.1). **General Identifiers**

- **state** (*Source: All*) – State name.

- **state_cd** (*Source: All*) – Two-letter state code.
- **year** (*Source: All*) – Reporting year.

CMS-Specific Variables

The CMS data are indexed by plan, program, and state; each row in the CMS data has a unique combination of these three variables. That is, one plan name can be paired with different programs/states, and one program name can be paired with different plans/states. CMS names have minor inconsistencies year-over-year and should not be used to join years of data.

- **cms_plan_name** – The name of the managed care plan (e.g., "Aetna Better Health").
- **cms_program_name** – The state-level program through which the managed care plan is offered, with the specific plan type in parentheses
- **cms_geographic_region** – The geographic area in which the plan operates (note: this variable is inconsistently formatted and should be used with caution).
- **cms_parent_organization** – The organization overseeing the plan at the state level (not the corporate parent). Generally equivalent to kff_mco_name, but may include typos or inconsistent names for the same organizations within states.
- **cms_medicaid_only_enrollment** – Number of plan enrollees who are not dually eligible for Medicaid and Medicare.
- **cms_dual_enrollment** – Number of plan enrollees who are dually eligible.
- **cms_total_enrollment** – Total number of plan enrollees.

TAF-Specific Variables

The TAF data are indexed by a combination of mc_plan_id and state. The same mc_plan_id may be used by multiple states. mc_plan_id is consistent across years within state and can be used to join multiple years of data, with caveats (see Appendix V).

- **taf_mc_plan_id** – Managed care plan identification number. Plan IDs are prefaced with an asterist ("*") to preserve leading zeros in ID numbers when the data is saved in .csv format.

- **taf_mc_plan_name** – Official contract name of the managed care plan or organization. Currently, we are leaving this variable out of the crosswalk on our public GitHub page until we have explicit permission from ResDac to include it.**taf_mc_plan_type_cd** – Plan type classification. Refer to the TAF Annual Plan Codebook for information on plan type code values. Currently, we are leaving this variable out of the crosswalk on our public GitHub page until we have explicit permission from ResDac to include it.
- **taf_enrollment** – Number of unique enrollees in the plan (not a point-in-time count; counts all enrollees at any time during the year).
- **taf_enrollment_in_claims** – Number of plan enrollees with at least one claim during the year.
- **taf_no_enrollment_or_claims –** The TAF data shows no one enrolled or making claims toward this plan. These plans are unlikely to have really been operating in the year of data in which they appear this way, but are include to facilitate joining the crosswalk and TAF managed care plan files.

KFF-Specific Variables

The KFF data are indexed by a combination of state and kff_mco_name. The same MCO name may be used by multiple states. kff_parent_firm is consistent across states and can be used to identify parent firms without using state. There are minor inconsistencies in kff_mco_name year-over-year, and it should not be used to join multiple years of data. KFF variables are only available for plans that are in the CMS data and that KFF considers comprehensive. Not all plans offered by comprehensive plans are included (See Appendix V).

- **kff_mco_name** – Name of the MCO, for comprehensive MCOs only.
- **kff_comprehensive_plan** – indicates KFF considers the plan in question to be a comprehensive managed care plan.

kff_parent_firm – Corporate owner of the MCO (listed by KFF if the company operates in at least two states)

CHSE-Created Variables

CHSE has added several variables to facilitate research, which are non-missing for all rows except where noted. state and chse_mco_name in combination uniquely identify MCOs. chse_mco_name is not consistent across states, but is consistent year-over-year within state. chse_parent_firm is consistent across all states and years, except in instances where a firm changed its name or was acquired by another firm (e.g. when Centene bought WellCare).

- **chse_mco_name** – Name of the MCO, for all MCO types
- **chse_parent_firm** – Corporate owner of the MCO
- **chse_blue_cross_blue_shield** – Flag indicating whether the plan is a Blue Cross Blue Shield (BCBS) plan. Multiple parent firms offer BCBS plans.
- **chse_matched_last_year** Flag indicating whether a plan appeared in both TAF and CMS data in the previous. Missing when CMS data unavailable in the previous year.
-
- **chse_matched_next_year** – Flag indicating whether a plan appeared in both TAF and CMS data in the subsequent year. Missing when CMS data unavailable in the subsequent year.

3. Using the Crosswalk

3.1 Choosing a Level of Aggregation

Plans are not a one-to-one match across different sources. For example, multiple CMS program names can match to a single TAF mc_plan_id. Therefore, it is often necessary to aggregate (deduplicate) based on the desired level of analysis. Researchers should use the following variables to analyze data at each level:

1. **Plan Level: state + cms_program_name + cms_plan_name for rows with CMS data available, otherwise state + taf_mc_plan_name.** This is the most granular level outside of TAF; each plan represents a separate option available to enrollees.
2. **MCO Level: state + kff_mco_name or chse_mco_name.** This aggregates plans into managed care organizations. A managed care organization may offer multiple plans within a state.
3. **Parent Firm Level: chse_parent_firm.** This will aggregate plans for which the MCO shares the same corporate owner, across states.

3.2 Resolving Conflicts

Discrepancies between sources can arise due to:

- **Missing plans** – Some plans appear in one dataset but not another due to reporting differences.
- **Classification differences** – Plan type according to CMS (based on program and plan name), from TAF (based on mc_plan_type_cd) and from KFF (all designated comprehensive) may conflict.

3.3 Joining with TAF data

To use the crosswalk with TAF data:

1. Extract mc_plan_id from TAF demographic, claims, or APL files.
2. Join the crosswalk to the TAF data at the mc_plan_id-state_cd level. Note that the crosswalk is not indexed by mc_plan_id and state_cd, as there are some instances where multiple CMS program or plan names or KFF MCO names match a single TAF mc_plan_id. For this reason, to conduct a one-to-one or one-to-many join with the TAF data, the crosswalk must be aggregated to the mc_plan_id-state_cd level first. To conduct a many-to-one join with the TAF data, the TAF data must first be indexed by mc_plan_id and state_cd.
3. Filter plans based on research needs (e.g., comprehensive vs. limited-benefit plans).
4. Use crosswalk variables for aggregation and analysis.

4. Limitations

This is a best-effort crosswalk; it may not resolve all inconsistencies between sources, and discrepancies may exist due to inconsistencies in data sources.

5. How to Cite This Crosswalk

To cite this crosswalk, use the following reference:

Hennessy, C. (2025). *Managed Care Crosswalks*. Center for Health Systems Effectiveness (CHSE), OHSU.

In addition, be sure to cite the TAF, KFF and CMS datasets as well.

6. Contact Information

For questions or comments, contact:

Conor Hennessy
Senior TAF Data Engineer
Center for Health Systems Effectiveness, OHSU
hennessc@ohsu.edu

7. References

U.S. Centers for Medicare & Medicaid Services. *Medicaid enrollment data by state, program, and managed care organization: 2017-2021*. U.S. Department of Health & Human Services. Retrieved February 6, 2025, from https://data.medicaid.gov/dataset/0bef7b8a-c663-5b14-9a46-0b5c2b86b0fe/data?conditions[0][resource]=t&conditions[0][property]=year&conditions[0][value]=2014&conditions[0][operator]=%3D

Kaiser Family Foundation. (n.d.). *Medicaid enrollment by MCO*. Kaiser Family Foundation. Retrieved February 6, 2025, from https://www.kff.org/other/state-indicator/medicaid-enrollment-by-mco/?

Kaiser Family Foundation. (n.d.). *Medicaid enrollment by MCO*. Kaiser Family Foundation. Retrieved February 6, 2025, from Internet Archive: https://web.archive.org/web/20240101000000/https://www.kff.org/other/state-indicator/medicaid-enrollment-by-mco/?currentTimeframe=0&sortModel=%7B%22colId%22:%22State%22,%22sort%22:%22asc%22%7D

TAF Data Access: ResDAC

U.S. Centers for Medicare & Medicaid Services. (n.d.). *TMSIS Analytical File (APL) data*. U.S. Department of Health & Human Services.

8. List of Appendices

Appendix I – Methodology for matching CMS and TAF Data

Appendix II: Reasons TAF and CMS plans might not match

Appendix III – Methodology for matching CMS and KFF Data

Appendix IV – Methodology for creating the 'CHSE' variables

Appendix V – Additional notes on the KFF Data

Appendix I – Methodology for matching CMS and TAF Data

The first step in creating the crosswalk was matching mc_plan_ids in the TAF data to program/plan combinations in the CMS data, state-by-state and year-by-year. To aid in the matching process, we calculated enrollment numbers for each TAF managed care plan, for comparison with CMS enrollment numbers. we found TAF enrollment numbers were higher, which is expected since TAF's enrollment definition includes everyone enrolled in any month during a year, whereas CMS provides point-in-time snapshots of enrollment.

Plans were matched using the following steps. This was executed manually (not by programming), so some interpretation and judgement was required.

Step 1: Match CMS and TAF plans that have similar names.

Step 2: Use enrollment numbers to confirm name matches.

- If CMS and TAF enrollment numbers are similar, the match is confirmed.
- If numbers are significantly different, check if the plans have closer enrollment numbers in other years. If so, the match is confirmed.

Step 3: Use enrollment numbers to match plans with dissimilar names.

- For the remaining plans, identify plans with similar enrollment numbers despite dissimilar names.
- Search the plans' cleaned names (with words like 'inc' or 'LLC' removed) on Google, using the following search string

  <CMS plan name> <TAF plan name> <state>

  Review the first ten results to see if they shed light on whether the two plans are the same. For example, in 2020 in Arizona the CMS plan "Magellan Complete Care" and the TAF plan "MOLINA COMPLETE CARE" have similar enrollments. Googling reveals that Molina bought the plan in 2020. TAF data reflects this change, while CMS's point-in-time analysis does not.

Step 4: Resolve cases where multiple CMS plans match multiple TAF plans.

Step 4a: Resolve multiple matches using mc_plan_id patterns.

In some states, multiple CMS plans match to multiple TAF plans. For example, each year in Texas, there are three CMS Amerigroup program/plan combinations – 'Amerigroup STAR', 'Amerigroup STAR PLUS', and 'Amerigroup STAR Kids' – but more than three TAF plans with names like 'Amerigroup.'

mc_plan_id patterns in combination with enrollment numbers can be used to determine which individual plans match. For example, in Texas, several mc_plan_ids start with the letter K. The sum of these plans' enrollment matches the enrollment in the 'STAR Kids' plans in TAF. This can be used to create matches at the individual plan level.

Step4b: Resolve multiple matches in California.

In California, sometimes one company offers different plans with similar names in different parts of the state. The bene_cnty_cd variable in the TAF demographic files can be used to determine the county where most enrollees live for each mc_plan_id, This can be compared with cms_geographic_region variable (which details were plans operate) to make individual plan level matches between TAF and CMS.

Step 4c: Resolve multiple matches in Florida.

In Florida, sometimes one company offers different plans for different populations (e.g. child welfare plans or MLTSS plans). Age data from the TAF demographic files can be used to determine the median age of enrollees for each mc_plan_id, and this can be used to distinguish between plan types.

Step 5: Note remaining inconsistencies.

These scenarios cannot be resolved with the available data:

- A single CMS plan matches two or more TAF plans, or a single TAF plan matches two or more CMS plans. Since it is not possible to break down the single plan further, these cases are left as is.
- Either the TAF plan or the CMS plan has no match in the other source. The data sources' plans may differ for several reasons (see next appendix), so it is expected some plans will not match. These cases are left as is.

Appendix II: Reasons TAF and CMS plans might not match

1. Year-long data vs point-in-time data

The TAF data comes from all twelve months of the year, whereas the CMS data is a snapshot from one point in time. As such, if a plan was only operating for part of the year, it may be included in the TAF data and not the CMS data. For this reason, we added the flags 'chse_matched_next_year' and 'chse_matched_last_year' to indicate that an unmatched plan was matched the year before or the year after, and the mismatch is likely a timing issue. However, these flags are insufficient to capture instances where a plan started

operating in 2021 (our current final year of data), ceased operating in 2017 (our first year of data) or only ever operated for a part of one year.

2. Differences in what is considered managed care

The TAF data includes many 'managed care plans' that are coded by TAF as 'Health/Medical homes' (mc_plan_type_cd == 70). None of these appear in the CMS data. This is likely not due to an omission, but because CMS does not consider these to be managed care plans.

3. Errors in the data

The TAF and CMS data contain errors and inconsistent use of name variables across states. Plans sometimes appear in years in which they are no longer operational, and in CMS they sometimes have a cms_program_name that is misleading. For example:

- In 2020 TAF data, there is a behavioral health managed care plan called 'Hudson River PC', a psychiatric center that has been closed since 2012. It only has one enrollee and no claims in TAF. This and some other discontinued plans can be easily filtered out by their enrollee and/or claim numbers.
- In CMS 2018 data, the plan name "Access Dental Plan, LLC" in Oregon has a program name denoting a comprehensive MCO, despite being a dental plan. Some erroneous comprehensive MCOs can be filtered out using KFF data. The KFF data is the closest thing to a gold standard that exists in the crosswalk but is only available for comprehensive MCOs, and should be used with caution as some plans without the flag could be legitimate MCOs (see #1 in this section).

4. Errors in the crosswalk

The crosswalk is subject to human judgement calls and could contain errors. If you think you have found an error in the crosswalk, please contact Conor Hennessy.

Appendix III – Methodology for matching CMS and KFF Data

We matched KFF MCO names to sets of CMS plans based on the names of the plans and MCOs, which tend to be similar. For 2018 onward (the first year KFF provides enrollment numbers by MCO), these matches were verified using enrollment numbers from each source.

Appendix IV – Methodology for creating the 'CHSE' variables

chse_parent_firm:

chse_parent_firm was derived from the kff_parent_firm field where available, otherwise from cms_parent_organization. In the latter cases, we corrected the field as needed by reviewing the first ten search results from the following Google searches:

"who owns <CMS and/or TAF plan name> <state>"
"<CMS and/or TAF plan name> parent firm <state>"

chse_blue_cross_blue_shield_flag:

Any plan that has 'blue', 'cross', or 'shield' in its plan name per TAF and/or CMS was googled to determine if it was a BCBS plan.

chse_mco_name:

Both cms_parent_organization and kff_mco_name are used to create this variable. Despite its name, cms_parent_organization is often the same entity as kff_mco_name (for comprehensive plans, where the latter is available), and appears in most cases to represent an MCO, not a parent organization, even if the MCO has a parent organization in reality. These are two examples of the relationship between cms_parent_organization and kff_mco_name:

- In 2021 in Oregon, 'Trillium Community Health Plan' and 'Trillium Community Health Plan - Tri-County' are two separate cms_parent_organizations. They are also two separate MCOs in the KFF data. The actual parent organization of both Trillium MCOs is Centene.
- In Missouri, United Health Care offers four plans (eastern, western, central, southwest) which all have 'United Healthcare' as the cms_parent_organization. In the KFF data, just one United Healthcare MCO appears.

Comparing the count of cms_parent_organizations to the count of kff_mco_names per state confirms that the former is usually an MCO. In most cases, the counts match. When

they do not, the discrepancy can be attributed to a likely mistake in the CMS field—for example, in Michigan 'Meridian Health Plan of Michigan' and 'Meridian Health Plan' are both listed as parent organizations. If corrected, the counts match. In other instances, the unmatched counts are explained by differences in which entities KFF considers to be an MCO; footnotes to KFF data list the states, but not specific instances, where this is true. Lastly, in Utah, the discrepancy is because KFF considers the 'HOME' MCO to be separate from 'Healthy U', despite both having the parent organization 'University of Utah Health Plans'. In all, the discrepant counts can be resolved for over 99% of cases.

Therefore, we use cms_parent_organization as a proxy for MCO when kff_mco_name is unavailable, with some additional steps. We must first check to see if the plan in question shares a cms_parent_organization with a plan that *does* have a kff_mco_name – if it does, it should be assigned that plan's kff_mco_name. Furthermore, due to the year-to-year inconsistencies in cms_parent_organization described above, we must manually fix some cms_parent_organizations.

Lastly, since TAF plans that do not match with CMS data lack a cms_parent_organization, for these plans we fill chse_mco_name with chse_parent_firm, except when the taf_mc_plan_name indicates the plan was likely offered by an MCO associated with similar plan names.

Finally, chse_mco_name is designed to allow merging of data across multiple years. Thus, the procedure described above was followed for 2021, but for prior years, the chse_mco_names were first assigned by merging the data with the following year's data using taf_mc_plan_id. Where a taf_mc_plan_id does not appear in the following years data, it is assigned a chse_mco_name according to the process described above.

<u>Pseudocode for creating chse_mco_name from other crosswalk variables:</u>

For the year 2021:

For each taf_mc_plan_id or CMS program/plan combination:

1. If there is a kff_mco_name:

    assign kff_mco_name value

2. Else if there is a cms_parent_organization:

a. If the same cms_parent_organization is paired with a kff_mco_name in other rows, assign that kff_mco_name*

b. Else assign cms_parent_organization

3. Else if the taf_mc_plan_name appears likely to belong to an MCO already in the data, assign that chse_mco_name*

4. Else assign chse_parent_firm

For years prior to 2021:

(Conducted in descending order by year)

Merge with the subsequent year's data using mc_plan_id and state_cd

1. If the merge results in a chse_mco_name, use that name

2. Else if no resulting name, check the plan and program names of CMS plans that did not match with TAF for similarities to CMS plans in the next year. If similar, assign that plan's chse_mco_name (because merging on mc_plan_id will skip unmatched CMS plans)*

3. Else assign kff_mco_name, cms_parent_organization_name, or chse_parent firm according to the procedure for 2021.

Appendix V – Additional notes on the KFF Data

Year-to-year changes

The KFF data was retrieved via the Internet Archive, also sometimes called the Wayback Machine. KFF, to the best of our knowledge, supplies on its website data from only the most recent year.

Our crosswalks incorporate KFF data from multiple years, and are designed for cross-year analysis, such that researchers can explore entries and exits of comprehensive MCOs from the market. However, because we use archived versions of pre-2021 KFF data, we do not know for certain that changes in the MCO landscape from year to year reflect an actual change versus a change in KFF's understanding of the landscape.

Using kff_mco_name to identify enrollees in comprehensive plans

The crosswalk matches comprehensive plan(s) in CMS to the appropriate MCO in KFF; these plans are assigned a kff_mco_name. Limited-benefit plans offered by the comprehensive MCO will not be matched. For researchers who wish to consider all plans offered by a comprehensive MCO, the chse_mco_name variable should be used rather than the kff_mco_name variable.

Appendix V – Joining multiple crosswalk years with taf_mc_plan_id

In general, mc_plan_ids are consistent for the same plans year-over-year, and can be used to merge multiple years of data. There are a small number of instances where an mc_plan_id was re-assigned despite no apparent changes to the managed care plan it referred to. However, in these instances, the plan, mco, and parent firm names of that mc_plan_id should remain consistent, meaning it is still possible to join multiple years of data on mc_plan_id, so long as the data is then aggregated to the plan, MCO, or firm level after the join.

Example 1 – note that all TAF data from '123' and '456' will be attributed to 'Health MCO' in all years, even if the years of data are joined on mc_plan_id

Year 1 MCO Name	Year 1 mc_plan_id	Year 2 MCO Name	Year 2 mc_plan_id	Year 3 MCO Name	Year 3 mc_plan_id
'Health MCO'	123	'Health MCO'	123	'Health MCO'	456
NA	NA	'Health MCO'	456	NA	NA

Example 2 – note that all TAF data from '123' and '456' will still be attributed to 'Health MCO' in all years, even though in no year do we see that '123' and '456' refer to the same plan.

Year 1 MCO Name	Year 1 mc_plan_id	Year 2 MCO Name	Year 2 mc_plan_id	Year 3 MCO Name	Year 3 mc_plan_id

‘Health MCO’	123	‘Health MCO’	123	‘Health MCO’	456