HCO THRIVE

CMS Quality Compare Aggregation
Technical Documentation: Methodology
Date: April 3, 2025
Version: 0.1

Contents

Overview

The purpose of this document is to describe HCOThrive's aggregation methodology for CMS' provider quality data (e.g., Hospital Compare, Home Health Agency Compare, and Nursing Home Compare).

We believe that longitudinal data had the potential to be useful in the hands of stakeholders interested in organizational analysis, public reporting, and research among other topics. Currently, there are a number of barriers that stand in the way of longitudinal analysis. In line with our mission to facilitate insight on healthcare organizations (HCOs), we currently provide both free and premium aggregated HCO data offerings for public use. Accordingly, this technical documentation provides a brief overview of encountered obstacles and the aggregation methodology used to overcome them.

CMS Quality-Compare Data

The following sections detail 1) common issues and challenges to most if not all CMS Quality Compare data as well as 2) provider-specific obstacles and pitfalls. Note that the provider data sections are limited to only those which required significant modification.

Common Issues

Variable Names

In many instances the Care Compare Provider Data available on CMS does not use variable names (or at least not by any standard definition) and instead provide full measure descriptions. This is not a major difficulty in and of itself, but the fact that these measure descriptions often change from period to period either in subtle ways such as a minor typo or more radically due to alternate wording.

Date-Range Correspondance

In some instances across providers, measure dates are only released as a separate file/dataset. This is not problematic in and of itself, however matching dates to measures (measure descriptions in this case) becomes a rather time consuming task due to the fact there rarely is an exact correspondence between said measure descriptions. More specifically, measure descriptions often do not match so a standard database-style join on measures requires manual intervention and research. Finally, there are cases where the date range needs to be calculated manually using the data documentation and other sources.

Documentation

Documentation for early releases is very limited and in many cases is not machine-readable. In some cases, documentation either lags ahead or behind of the actual data release. It should be noted that these issues are found primarily among CMS' earlier quality data releases and that the overall documentation quality tends to improve dramatically over time.

Aggregation Availability

CMS public data release mandate requires the past seven (7) years of archive quality data to be made available to the public. Releases prior to this can be still be accessed via the "Wayback-Machine" (Archive, n.d.), but this is not always a straightforward proposition.

Standard Remedies

In summary, we, at a minimum, perform the following procedures,

1. Identify all input files with relevant meta information (source, date, etc.)
2. Create machine AND human-readable variable names
3. Ensure all recorded values correspond to a date range
4. Remove all `null`, missing, or "N/A" values
5. Check and remove all duplicate values
6. Provide acceptable documentation

Hospital-Compare (HC)

HC-Readmissions-Complications-Deaths (RCD)

This data is currently a combination of two different datasets available on Hospital Compare: 1) readmissions/unplanned visits and 2) complications and deaths. The reason for this is that over the years readmissions (R), complications (C), and deaths (D) have been released in various combinations (e.g., `Complications and Deaths - Hospital.csv`, `Readmissions and Deaths - Hospital.csv`, `Readmissions Complications and Deaths - Hospital.csv`, etc.) (Medicare & Medicaid Services, 2024a). Thus, for the purpose of aggregation, these various measures have been combined to facilitate historical analysis.

HC-Verterans-Administration (VA)

Veteran Health Administration quality data which currently includes timely and effective care and inpatient psychiatric facility (IPF) VA hospital data files (Medicare & Medicaid Services, 2024f). This dataset also includes past VA datasets including patient safety indicators, patient experience of care measures, and patient safety indicators. Note that all years and variables are included in the Basic plan.

Home Health Compare (HHC)

HHC-Provider

The only major alteration we undertook for HHC-Provider Data was to calculate a custom date range for the Quality of Patient Care Star Rating, which is not provided as part of either the separate Provider or HHCAHPS date file. Given that the QPC Star Rating is an aggregate of multiple (7) measures (Timely Initiation of Care, Improvement in Ambulation, etc.), this range was simply calculated as having the minimum start date of all composite measures and the maximum end date of all composite measures (Medicare & Medicaid Services, 2024b).

Nursing Home Compare (NHC)

As part of the available NHC data, CMS releases a "provider" dataset (Medicare & Medicaid Services, 2024c) that contains monthly star ratings and summaries of the variables that are used in calculating the ratings. This data has a large number of variables with varying data ranges which can complicate analysis for the end user. To address this, we've separated out several key groupings into their own distinct datasets: 1) ratings and 2) staffing ratios/turnover.

NHC-Provider-Ratings

For user convenience ratings are both presented on both a monthly and quarterly basis whereas only the current month is actively displayed on NHC at a time. Quarterly ratings calculate the average rating for the quarter to-date. So, for example, the quarterly rating as of January corresponds directly to the monthly rating for January, the quarterly rating of January and February is the average of the two months,

and so on. It is important to note that that a particular monthly rating may change post-hoc due to new information being added to the nursing home Minimum Data Set (MDS) which is the main NHC measure repository (Medicare & Medicaid Services, 2024d). In the case where a value for a previous month changes, we use the most recent rating replacing old ratings with newer ones if/when they change.

NHC-Provider-Payroll-Based-Journal (PBJ)

This dataset is separated out to only include PBJ specific variables. Note that only relatively recently has CMS begun including PBJ variable date-ranges with each release. For prior releases, we calculated the date range retrospectively based on the PBJ documentation (Medicare & Medicaid Services, 2024e). Also note that CMS began using PBJ over the Certification and Survey Provider Enhanced Reports (CASPER) system as of April, 2018. As of this time, we do not offer CASPER data due to issues with determining measure dates although this may change in the future.

References

Archive, T. I. (n.d.). *Internet archive: Wayback machine*. https://archive.org/web/

Medicare & Medicaid Services, C. for. (2024a). *About the data*. https://data.cms.gov/provider-data/topics/hospitals/about-data

Medicare & Medicaid Services, C. for. (2024b). *About the data*. https://data.cms.gov/provider-data/topics/home-health-services/about-the-data

Medicare & Medicaid Services, C. for. (2024c). *Data sources*. https://data.cms.gov/provider-data/topics/nursing-homes

Medicare & Medicaid Services, C. for. (2024d). *Nursing home quality initiative*. https://www.cms.gov/medicare/quality/nursing-home-improvement

Medicare & Medicaid Services, C. for. (2024e). *Staffing data submission payroll based journal (PBJ)*. https://www.cms.gov/medicare/quality/nursing-home-improvement/staffing-data-submission

Medicare & Medicaid Services, C. for. (2024f). *Veterans health administration (VHA) hospitals*. https://data.cms.gov/provider-data/topics/hospitals/veterans-administration-hospitals