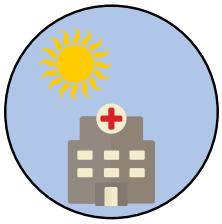

HCO THRIVE

HCRIS 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 provider financial data.

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.

HCRIS Aggregate Financial Data

The source of provider aggregate financial data is the Healthcare Cost Report Information System (HCRIS) (Medicare & Medicaid Services, 2023a).

Healthcare Cost Reporting Information System (HCRIS)

The CMS HCRIS data consists of cost reports submitted by providers in order to determine Medicare and Medicaid reimbursement and rule-making. Cost reports consist of multiple alphanumeric worksheets that record utilization, charges, and, of course, costs.

> "Medicare-certified institutional providers are required to submit an annual cost report to a Medicare Administrative Contractor (MAC). The cost report contains provider information such as facility characteristics, utilization data, cost and charges by cost center (in total and for Medicare), Medicare settlement data, and financial statement data. CMS maintains the cost report data in the Healthcare Provider Cost Reporting Information System (HCRIS)" (Medicare & Medicaid Services, 2023a).

Data-Specific Challenges

Ease-of-Use

This particular heading primarily refers to the overall complexity and breadth of the CMS HCRIS data. Data is stored in multiple different files and across different cost-report versions depending on the release date. Within a cost-report, there are typically 12-15+ total worksheets with anywhere from 30 to 100+ potential line numbers and entries (Medicare & Medicaid Services, n.d.). Form instruction manuals are long, dense, and tersely written. Although CMS has begun releasing a PUF file (Medicare & Medicaid Services, 2023b) with actual variable names, this is limited in terms of scope and only dates back several years meaning in all but the most basic cases, the user must transform and calculate desired indicators themselves. Combined, these factors contribute to a relatively onerous transformation process although the mostly systematic nature of the releases is certainly a positive.

	RPT_REC_NUM	WKSHT_CD	LINE_NUM	CLMN_NUM	ITM_VAL_NUM
1	268	A000000	00100	00200	1.150680e+05
2	268	A000000	00100	00300	1.150680e+05
3	268	A000000	00100	00400	1.878440e+05
4	268	A000000	00100	00500	3.029120e+05
5	268	A000000	00100	00600	-1.213000e+03
6	268	A000000	00100	00700	3.016990e+05
7	268	A000000	00200	00200	9.535700e+04
8	268	A000000	00200	00300	9.535700e+04
9	268	A000000	00200	00400	1.087100e+04
10	268	A000000	00200	00500	1.062280e+05

Figure 1: Raw HCRIS Data

Aggregation

HCRIS instructions allow providers to subscript certain lines on the cost-report in order to provide more detail (e.g.,

$$\text{Line } 25^{\text{NEW}} = \sum 25.01, 25.02, \ldots$$

) (Medicare & Medicaid Services, n.d.). Typically, this is done in order to enumerate subproviders, spread values across multiple time periods and/or programs, and list provider-specific allowable costs. In order to create a usable dataset, these subscripted line numbers need to be aggregated (summed) by report and provider. There are also partial line numbers (e.g., 20.10) that actually comprise part of the report and should not aggregated. Although CMS does provide roll-up files with their raw data releases, these do not include all worksheet lines and may be unreliable for some form versions (Medicare & Medicaid Services, 2023b).

Cost-Center Coding

Among the HCRIS Cost Report Forms, here are certain worksheets (commonly Worksheet A-D) where providers either have the option or are required to employ cost-center coding. As stated in the HCRIS documentation "*cost center coding is a methodology for standardizing the meaning of cost center labels as used by health care providers on the Medicare cost reports*" (Medicare & Medicaid Services, n.d.). What this means in practice is that a line number such as 12.01 may actually need to be aggregated under another line number/cost-center such as 15.

Form Correspondence

Over time, due to changing legislation and regulatory requirements, HCRIS forms are superseded by newer versions. Most of the time, changes between forms are significant with worksheets and line numbers both being replaced or removed. Ensuring maximum correspondence between between different forms often requires a detailed review of both form's separate documentation as well as any crosswalks or additional documents CMS provides.

Update/Replace Reports

The HCRIS system is updated quarterly for each relevant fiscal year. Updates continue for 2-3 years after the actual file fiscal year is over. This means that in order to stay up-to-date, multiple years of zip files must be processed each quarter. Additionally, each quarterly update will contain a number of updated reports for a small subset of providers. Here, it is important to note that each report number is unique meaning that the old reports are replaced and removed from the zip files upon an update.

Standard Remedies

As part of our standard aggregation methodology, 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

Dataset-Specific Remedies

Our financial data is intended to represent a significant enhancement in terms of ease-of-use for reporting and analysis purposes. Beyond the standard remedies referenced above, we also perform the following,

1. Roll-up and aggregate subscripted entries across all reports
2. Create longitudinal data-set across different forms
3. Replace data for partial or incomplete Cost Reports

References

Medicare & Medicaid Services, C. for. (2023a). *Cost reports*. https://www.cms.gov/research-statistics-data-and-systems/downloadable-public-use-files/cost-reports

Medicare & Medicaid Services, C. for. (2023b). *Hospital provider cost report*. https://data.cms.gov/provider-compliance/cost-report/hospital-provider-cost-report

Medicare & Medicaid Services, C. for. (n.d.). *The provider reimbursement manual - part 2*. https://www.cms.gov/regulations-and-guidance/guidance/manuals/paper-based-manuals-items/cms021935