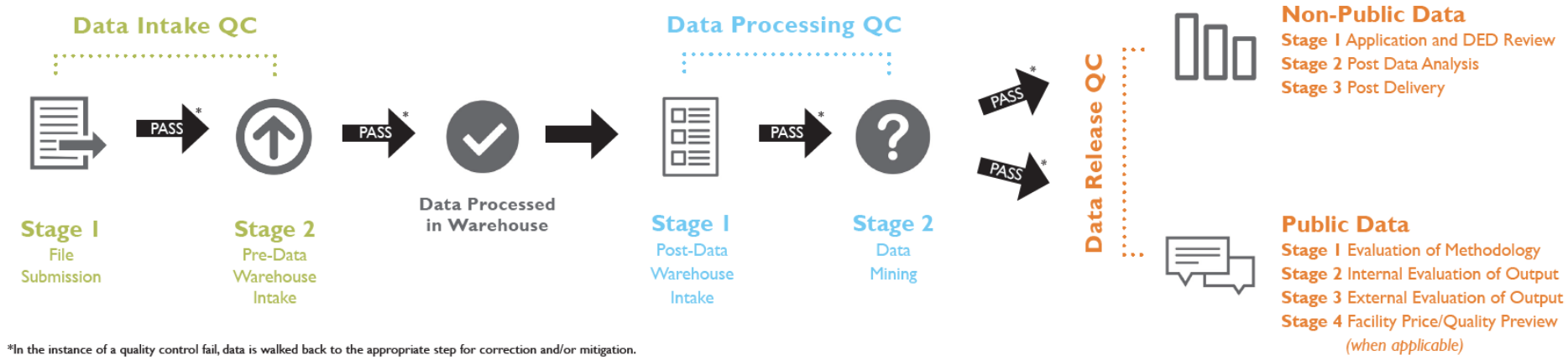
Colorado All Payer Claims Database

Data Quality Process: Intake, Processing, Release

The Colorado All Payer Claims Database (CO APCD) is the state's most comprehensive source of health care insurance claims information representing the majority of covered lives in the state across commercial health insurance plans, Medicare, and Medicaid. On a monthly basis, over 70 individual submitters across 33 major commercial health insurance payers and Medicaid submit claims data for their eligible members to the CO APCD. Payers submit claims information for all eligible members (both health care users and healthy non-users), representing over 4.5 million claims collected and processed monthly. Additionally, the Centers for Medicare & Medicaid Services (CMS) submit separate files with Medicare Fee-for-Service claims both quarterly and annually[1].

As administrator of the CO APCD, Center for Improving Value in Health Care (CIVHC) works with data warehouse managers Human Services Research Institute (HSRI) and NORC at the University of Chicago to ensure quality checks occur in multiple stages throughout the lifecycle of the data: upon intake, processing, and data analysis and release.

Management of each stage in the cycle is complex and multi-faceted. This document outlines the quality checks and processes that are continuously refined and improved to ensure data in the CO APCD is as complete and accurate as possible. For more details regarding our quality checks, please contact us at ColoradoAPCD@civhc.org.



*In the instance of a quality control fail, data is walked back to the appropriate step for correction and/or mitigation.

[1] CIVHC receives Centers for Medicare & Medicaid Fee-For-Service Medical data quarterly and Part D annually. It is important to note that Medicare FFS files do not come in the same format as claims submitted by commercial payers and Medicaid. Because of the discrepancy in file types, it takes additional time to process and intake Medicare claims and additional checks are conducted upon processing these claims.

DATA INTAKE

Stage 1: File Submission

After health insurance payers submit claims through the CO APCD Portal, the files are loaded to staging tables and evaluated using several hundred validation rules established to identify issues with the file layout and common data submission errors. A current list of validation rules is available to payers in the CO APCD Portal to assist them with understanding how their submissions are evaluated and to help minimize errors. The validation rules are reviewed and revised at least annually and updated as new issues are identified.

At a high level, the system automatically searches the files for common issues including:

1) **Failure-level issues:** Intrinsic data format issues that cannot be overridden, requiring a payer fix and resubmission.
2) **Structural-level issues:** File formatting errors requiring a payer fix and resubmission.
3) **Threshold-level issues:** Volume thresholds that estimate items such as the types and amount of claims that should be present for each individual payer. If the data submitted does not meet expectations, the payer may have to resubmit the data depending on the reason for the discrepancy.

Stage 2: Pre-Data Warehouse Intake

The pre-data warehouse intake stage assesses quality of the data after file submission validations in Stage 1 are complete, and calculated variables, member ID and provider processing has occurred. This stage evaluates whether any of the newly available files are resubmissions of data that were formerly part of the data warehouse.

Stage 2 also includes an assessment of demographics, claim volume trends, Per Member/Per Month (PMPM), member eligibility, and match of claims to eligibility. Gaps, deficits, and surpluses in these totals in one or more months could indicate submission issues:

1) **Demographics:** Key member demographic variables, such as gender and age, are examined for each payer and compared with previous values.
2) **Overall Claim Volume**: Overall claim and claim line volumes for the three types of claims and eligibility (dental, medical, and pharmacy) are evaluated for consistency and trend expectations.
3) **Payer Claim Volume:** Claim volume is tracked over time by submitter (i.e. Aetna) and submitter subgroups (i.e. Aetna's HMO submissions). Volume is examined by both paid and incurred date to ensure that records are complete, and volume by type of claims (for example, inpatient and outpatient) are compared to historic volumes for consistency.
4) **Dollar Volume:** Dollar volume is tracked over time by submitter and submitter subgroups. The dollar volume is compared over time to an expected dollar volume is received.
5) **Monthly Eligibility and Member Match to Eligibility:** Continuous eligibility and the number of eligible members in the eligibility file is compared to the number of members with claims. If the ratio of unique members in the claim file to unique members in the eligibility file for a payer is greater than 0.8 or the appropriate amount historically, the payer may have a gap in the eligibility file. If a member claim to eligibility ratio is less than 0.1, the payer may have incomplete claims for its members.
6) **Per Member/Per Month (PMPM):** For each type of claim (medical, dental, and pharmacy), for each service month included in the submission, a PMPM amount is calculated. PMPM amounts should be fairly consistent on a month-to-month basis, with any increases or decreases following a logical trend. Months within the submission period that have drastic jumps in PMPM amounts are investigated to determine if they are indicators of partial or duplicate submissions.

CIVHC and HSRI have Pass/Fail criteria for all Stage 2 checks which are used to determine whether a file passes this quality assurance stage. As with Stage 1, quality control checks in Stage 2 are reviewed and revised at least annually and as new issues are identified.

DATA PROCESSING

Stage 1: Post-Data Warehouse Intake

After files pass Stage 2 quality control in the intake process outlined above, they are moved into the CO APCD data warehouse. At this point, additional value-add work such as 3rd party groupers is conducted and the Stage 3 quality control is conducted to evaluate the completeness of the index/directory information and the value-add component outputs. This stage of quality control is conducted quarterly.

1) **Master Person Index:** For each month in the quarter being assessed, a percentage value is calculated that shows the percentage of persons in that month associated with a pre-existing person on the Master Person Index. Low percentage match rates can indicate problems with the identifiers that data submitters are providing or the process by which unique identifiers are being assigned. Low match rates can also show data submission gaps.
2) **Master Provider Directory:** For each month in the quarter being assessed, a percentage value is calculated which shows the percentage of providers in that month associated with a pre-existing provider in the Master Provider Directory. Provider information is also compared against third party information, such as the National Plan & Provider Enumeration System (NPPES), where available, in order to detect whether providers are being matched appropriately. Low percentage match rates can indicate problems with the identifiers that data submitters are providing or the process by which composite IDs are being assigned. Low match rates can also show data submission gaps.
3) **Value Add Components**: Each value add component, such as 3rd party groupers, has a unique quality control procedure. Value add components are not available for data analysis and release until QC procedures have been completed and documented.

Established Pass/Fail criteria are used to determine if data passes this assessment. If the criteria isn't met, the CIVHC/HSRI team perform edits on a given file or files to resolve one or more data quality issues. These changes are performed, documented, and tracked according to Data Anomaly Remediation Standard Operating Procedures. Once these changes are implemented, the data quality assessment is performed again and the passing percentages updated. The quality control checks outlined in this section are reviewed and revised at least annually and as new issues are identified.

Stage 2: Data Mining

A key component of the data quality process is proactive data mining. Data mining entails the review of data at all stages of the QC process outlined above to find previously undetected issues, develop new data quality checks and reports based on these issues, and develop action plans to resolve newly identified issues. Data mining goals are defined depending on issues identified or questions that arise from the internal CIVHC/HSRI/NORC analytic team or external stakeholders. Examples of issues/questions include:

- A data submitter has reported an issue in their data and it is necessary to examine the overall impact of that issue, and whether it effects other data submitters or the database as a whole
- A stakeholder has requested a custom report with data elements that have not been produced before and we need to understand if we have the underlying data to support the analysis

DATA RELEASE

CIVHC transforms data that has been validated through the stages above into actionable information for stakeholders through both public and non-public releases of information. Quality checks are incorporated into the data release process to ensure the information is as accurate as possible after performing additional analysis.

Non-Public Data Release

CIVHC provides custom analyses to requestors seeking to inform projects that advance the Triple Aim (better health, better care and lower costs). Once an application has been approved by CIVHC and the Data Release Review Committee, the application for data moves into the production phase. Multiple quality checks are initiated during the process of delivery to ensure the data provided is as accurate as possible.

Stage 1: Application and Data Element Dictionary Documentation Review - Thorough review of the application and associated Data Element Dictionary (DED) filters and tabs to ensure they align and meet the intent of the request.

Stage 2: Post Data Analysis

a. Confirm selections on DED matches output
b. Confirm row counts on all tables appear reasonable based on available information
c. Confirm all DED filters are properly applied to claims tables and eligibility table
d. Spot check issues that could have occurred due to complicated filter applications
e. Verify that output matches expected results (when external sources are available or results have an anticipated range of output)

Stage 3: Post Delivery - If any data issues or concerns are discovered by a client, they are investigated, added to an issue tracker, addressed in the short term to meet client needs, and fixed in the appropriate stage above to prevent reoccurrence moving forward.

Public Data Release

CIVHC provides publicly available data to help stakeholders identify ways to advance the Triple Aim (better health, better care and lower costs) and to help consumers make informed health care decisions. The CO APCD Advisory Committee helps CIVHC determine the measures and topics to provide publicly. Multiple quality checks are initiated during the process of public release to ensure the data provided is as accurate as possible.

Stage 1: Evaluation of Methodology – Analyst review of accuracy of defined methodology in software utilized

Stage 2: Internal Evaluation of Output – Analyst review of data output and verification against similar state and national data sources, when available, to confirm output is in alignment with expected results or similar analyses.

Stage 3: External Evaluation of Output – Stakeholder review of data is offered prior to release of public analysis to pro-actively identify any potential issues and answer questions related to methodology or other details impacting the analysis.

Stage 4: Facility Price/Quality Preview (when applicable) – Prior to the release of named facility price and quality information, CIVHC conducts a preview period with facilities to validate data that will be made publicly available.

The quality control and quality assurance processes identified in this document summarize more detailed information available by contacting ColoradoAPCD@civhc.org.