

Kentucky Health Data Trust Initiative

Principles and Techniques for Data Quality Reporting and Data Validation

Deliverables 4.1.4 and 4.3.2

Prepared for the Kentucky Health Data Trust
Interagency Governance Workgroup

Freedman HealthCare
July 7, 2015



Introduction

This white paper describes the various processes that state APCD teams can implement to check and validate data quality during the intake, management, and release cycles. Rigorous quality checks at all stages will help ensure that the data pulled into, and derived from, the APCD are organized, reliable, and accurate.

Techniques for Ensuring Data Quality at Intake

As submitters send data to the APCD, the APCD team should implement a preliminary round of format and credibility checks to test the quality of data at intake. The purpose of these quality checks is to verify that incoming data files fully conform to the APCD's submission specifications. States should implement the following techniques to check data quality at intake:

1. *Create data collection standards.* States should clearly define these submission specifications for all data submitters prior to implementing the APCD. These standards are often called the "data submission guide," and are found in both mandatory and voluntary APCD models. States with voluntary data submission can include their data collection standards as part of the data use agreements they execute with commercial payers. [Colorado's APCD](#) provides a good example of a data submission guide for mandatory filings.
2. *Build automated processes to check the data files.* The entity responsible for data intake should create automated tests to check and accept all incoming data files. These automated processes check all data elements against the technical specifications outlined in the data submission guide, as well as check other "mechanical" items such as non-standard codes, outlier codes, consistency in Start and End dates, etc. Furthermore, the tests check for reasonability of the data by comparing the current month's data file to that from the previous month as well as from twelve months' prior.

This suite of automated data quality checks should utilize a pass/fail rubric that the intake team creates to evaluate the severity of missing or misformatted data elements. After the data pass through the automated tests, the system will generate a pass/fail score and summary report for each data submission. An analyst from the data intake team should then review the report and communicate the findings to the data submitter. Figure 1 provides an example of an intake quality report for the Minnesota APCD.

Figure 1: Sample Quality Report at Intake, Minnesota APCD

MTIME	ELEMENT	COLNAME	VALID	INVALID	NULLCOUNT	THRESH	PAYERTHRESH	PRC	RESULT
201301	ME001	NAIC	10.0	0.0	0.0	100%		100.0	PASS
201301	ME002	NPLAN	8.0	2.0	0.0	0%		80.0	PASS
201301	ME003	PRODUCT	10.0	0.0	0.0	100%		100.0	PASS
201301	ME006	IGROUP	10.0	0.0	0.0	99.5%		100.0	PASS
201301	ME007	XPLAN	10.0	0.0	0.0	99.5%		100.0	PASS
201301	ME008	ESSN	10.0	0.0	0.0	0%		100.0	PASS
201301	ME008&ME009	ESSN_CONTRACT	10.0	0.0	0.0	100%		100.0	PASS
201301	ME009	CONTRACT	10.0	0.0	0.0	0%		100.0	PASS
201301	ME010	SEQNO	10.0	0.0	0.0	90%		100.0	PASS
201301	ME011	MEMSSN	10.0	0.0	0.0	0%		100.0	PASS
201301	ME012	REL	10.0	0.0	0.0	100%		100.0	PASS
201301	ME013	SEX	10.0	0.0	0.0	100%		100.0	PASS
201301	ME014	DOB	10.0	0.0	0.0	99.5%		100.0	PASS

Citation: "Minnesota Health Care Claims Reporting System." Presentation to the MN APCD Stakeholder Workgroup by Onpoint Health Data and the Minnesota Department of Health, September 30, 2014. Available at: <http://www.health.state.mn.us/healthreform/allpayer/apcdonpointpresentation093014.pdf>. Accessed on July 2, 2015.

3. *Identify and resolve data issues with the submitter as needed.* If the initial data submission failed the quality checks at intake, an analyst from the data intake team should alert the submitter and assist with problem resolution. States with mandatory data collection require data submitters to review, correct, and resubmit the data until all files pass the intake quality checks. The data intake team is responsible for monitoring and tracking all data submissions to demonstrate compliance. For cases in which a submitter is unable to provide a particular type or amount of data as specified in the submission guide, the intake team should develop an exemptions process. Rhode Island's [Technical Specifications Manual](#) for data submitters offers an example of this exemptions process (p. 50).

Voluntarily submitted data files are less likely to be standardized or resubmitted for corrections. In these cases, the data intake team should document all data gaps or issues so that downstream data users better understand any limitations of the data.

Finally, APCD intake quality checking processes need to evolve. Data submitters' processes change and fluctuate, sometimes from month to month. Experts within the submitting organization may leave, or legacy systems create limitations for data submission. To ensure consistency in data quality and flow, the APCD data intake team should maintain close communication with all data submitters. Ongoing meetings between the APCD data intake team and the data submitters are helpful for providing technical assistance and understanding necessary changes or corrections.

As an additional step to ensuring data quality, the APCD intake team should share their data checking criteria with all data submitters so that submitters can pre-check the files before sending them. This can help reduce administrative costs associated with data submissions for both the submitter and the data intake team, and can allow submitters to maintain a standard production schedule.

Techniques for Validating the Data Mart

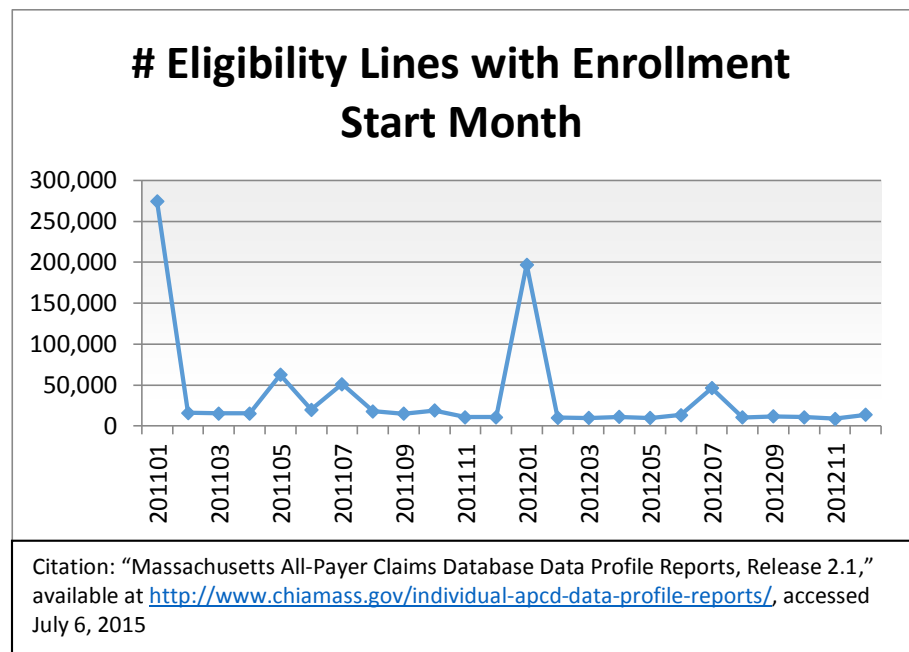
After the data submission files complete the initial quality checks and move into the data mart, the resultant files should pass through additional validation processes before entering production. These validation processes should measure the continuity and consistency of the data files month over month to verify trends internal to each file, as well as across all files. In addition, validation should compare measurements across providers to ensure that trends are consistent and steady over time. APCD technical teams should implement the following techniques for data validation:

1. *Check all files within a data submission for completeness.* Validation checks should examine all files within a carrier's data submission to confirm that every member listed in a claims, behavioral health, or pharmacy file has a corresponding member record. This will allow the data team to align individual members across all files and track them over time.
2. *Compare each data submission to that from the previous month and year for consistency in overall trends.* During validation, the APCD team should test the consistency of the compiled information across all months in the reporting period for each data submitter. Examples of validation processes include month-by-month member counts, claims lines per member per service month, per member per date of service month and per paid month, member and provider zip code distributions, and procedure and diagnosis frequencies. Periodic checks of overall membership and claims numbers can also help flag any significant or unexplained increases or decreases that would skew analyses. States should keep in mind that these validation checks are not designed to show meaningful variation in spending or utilization;

rather, their purpose is to ensure that the data are reasonable and show no major anomalies or discrepancies that would skew the overall trends.

Figure 2 is an example of one type of monitoring that the Massachusetts Center for Health Information and Analysis (the state's APCD administrator) prepared and presented on its website to provide transparency about the quality of the data as well as to show the total number of member records in each company's file.

Figure 2: Sample Enrollment Monitoring



3. *Benchmark the data against historical files as well as national and state data sources.*

The APCD team should check data files from each submitter against submissions from prior reporting periods to ensure that there have been no major spikes that could indicate an error in the data. The team should also benchmark the data against external data sources such as NAIC filings and state agency headcounts, to confirm that the numbers align with what is already known in the current universe.

The state's APCD team should include technical/programming staff with knowledge of claims data who can create trend files and baseline files to benchmark and compare data. From these validation checks, the APCD team can produce quality reports for each submitter that detail the results of the data quality and validation checks. A good example of this is the publicly released [Data Profile Reports](#) that the Massachusetts APCD produces for each carrier's medical claims and member eligibility submissions.

4. *Focus on validating data from comprehensive medical coverage plans.* While APCD data submission specifications typically cast a wide net by requiring submitters to report on all their lines of business, the APCD data team should focus its validation efforts on testing data from comprehensive medical coverage plans. Unlike plans that offer partial or limited coverage (e.g. Medigap, accident, or disability coverage), comprehensive medical coverage plans better represent the complete cost of care and are therefore most useful for cost analyses. In addition, data teams can more easily assess the completeness of the data in comprehensive coverage plans and reconcile them to other data submitted by carriers. Consequently, the most rapid and effective method for validating data quality and completeness is to test these comprehensive medical coverage plans.
5. *Normalize the data to account for differences that might skew the results.* The number of business days per month varies; some months are shorter or have more potential business days. Consequently, per member per month costs are typically higher during the months of March, August and October even when utilization rates are stable. Data analysts should normalize spending and utilization to business

days per month (PMPD or utilization per business day); otherwise, numbers become skewed by long months with no holidays (e.g. March, August, and October).

6. *Convene carrier staff to discuss and confirm member counts.* Staff from the APCD's business unit should meet concurrently with the insurance carrier's actuarial staff (in charge of annual rate review filings to the insurance regulator, among other reports) and APCD reporting staff (in charge of data submissions to the APCD) to compare their methods for reporting membership data. These two departments often work separately within the company and can have very little interaction with each other; in some states, this has led to discrepancies between APCD data and regulatory filings. To better align data submission methodologies across these two teams, the state APCD team should meet with the carrier's key financial and APCD reporting staff to discuss their data submission processes and agree on a common definition for data elements such as plan type (e.g., HMO, PPO, POS, HDHP, and Indemnity) and product type (e.g. medical/health, ancillary drug, dental, cancer, hospital, other). Aligning these definitions will help the carrier and APCD team confirm member counts by ensuring that members have been assigned to the correct plan or product "bucket" across all data submissions. The APCD team should meet with these two teams during the initial APCD build, as well as on an annual basis to compare summary APCD reports to the carrier's annual rate filings.
7. *Set expectations for quality compliance with mandatory data collection.* Data validation is a critical component of the APCD process, and is especially necessary for states with mandatory data collection. In these cases, monitoring carriers' submissions is a constant and ongoing function, and those within the APCD team who perform this function play a compliance role on behalf of the state. States with mandatory data collection have found it helpful to establish penalties for noncompliance as part of the APCD regulations, as this provides a standardized enforcement framework and sets clear expectations for all data submitters. However, because APCD teams rely on close collaboration and communication with carriers to address data submission issues as they arise, penalties are rarely applied and should not be considered as a potential revenue source for the APCD. Table 1 provides an example of how an APCD team tracks a data submission through the various quality checks in the Rhode Island APCD.

Table 1: Tracking a Data Submission in the Rhode Island APCD

Stage	Status	Description	Typical Follow-Up Action Required
PRELIM	REJECTED	File has been rejected in the preliminary stage since a preceding version has been extracted	Reason for resubmission
PRELIM	FAIL	File has failed the preliminary stage for not meeting field requirements	Resubmission
LOAD	FAIL	File has failed the load stage for not meeting the default threshold on particular fields	Resubmission or request for a waiver to the threshold
DELETE	DONE	File has been replaced and deleted	None
DQ	FAIL	File has failed the data quality validations	Resubmission to correct failing validations
DQ	HOLD	File has some questionable data quality validations that are failing	Resubmission or explanation on failing validations
DQ	REVIEW	File has entered data quality review	Manual review by Onpoint's staff
DQ	PASS	File has passed the data quality validations	None
REPLACED	FAIL	Failed file has been replaced	None
REPLACED	PASS	Passed file has been replaced	None
TRANSMIT	INHOUSE	File has been received in house and is in the queue for processing	None

Citation: Technical Specifications Manual: Rhode Island's All-Payer Claims Database (APCD), Version 1.4. Table 3: "Data Stage and Status Categories." (March 2015) Available at:

<http://www.health.ri.gov/materialbyothers/RIAllPayerClaimsDatabaseTechnicalSpecificationsManual.pdf>. Accessed on July 2, 2015.

Techniques for Validating Data Extracts and BI Tools

As the data move into production within the data mart, the APCD team should implement additional validation measures to maintain the accuracy and integrity of all data retrieved for reports and ad-hoc customer services. At this stage, it is also the responsibility of the APCD team to provide sufficient training and communication to data users so that they are correctly using and interpreting the data. The state should put sufficient data governance processes in place to ensure that any information pulled from the APCD is accurately reported.

1. *Designate agency-specific data access leads.* States should engage various state agencies to access and use the data as trusted partners; however, states should also take steps to control dissemination of the data. After identifying the designated agency users, the APCD data manager (whether it be the state administrator or an external vendor) should create secure access to the data and set up role-based permissions for each user. Some APCD states, such as Rhode Island, designate a "power user" at each participating state agency to manage these permissions and ensure data quality.
2. *Ensure that designated users know how to access and interpret the data correctly.* The APCD data manager should provide hands-on assistance and coaching to all users to drive their expanded use of the data. To ensure that users understand the data and know how to appropriately access and apply the information, the data manager should create user guides and documentation; in addition, the APCD team should provide training and ongoing technical assistance to users to help them best leverage the information and troubleshoot issues. States that are just beginning to use their APCDs can engage their

state agency users to act as “beta testers” to flag issues in the data mart and BI tool that need resolution or improvement.

3. *Publish clear methodologies describing the development of the database.* The APCD data manager should produce documentation that clearly describes the quality of the data in the data mart, including but not limited to the following elements: what the database does and does not include, the limitations of using the most recent data, and that which a user might or might not be able to learn from the data. In addition, these methodologies should clearly define any restrictions or parameters that may exist for publication, such as cell sizes, minimum number of payers, and median price reports. This documentation should be easily accessible to all users and potential audiences, to help prevent any misinterpretation or misuse of APCD information. The New Hampshire Insurance Department, which oversees the NH APCD, engaged an outside evaluator to analyze the quality of the APCD; the state subsequently released a [public report](#) of the findings in 2012. **Table 2** is an excerpt from a public use data file that the Oregon APCD released, detailing the layout and data elements included in the APCD’s medical claims data.

	A	B	C	D	E	F
1	Data Elements					
2	Name	Long Name	SAS type	SQL type	Length	Notes
3	year	Calendar year	numeric	int	4	
4	patid	Encrypted unique patient identifier	text	varchar	30	
5	gender	Gender	text	char	1	Excluded if unknown
6	agegrp	5-year age group	text	varchar	5	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54,
7	race	Race	numeric	int	1	See Lookup Table LT race
8	ethnicity	Ethnicity	numeric	int	1	See Lookup Table LT eth
9	language	Primary spoken language	text	varchar	3	Code set available from the NISO web site
10	metro	Metropolitan indicator	numeric	int	1	1 if metropolitan MSA; 0 if non-MSA area; blank if missing
11	paid	Total payment	dollar	money	12	
12	patpaid	Patient pay amount	dollar	money	12	
13	paycat	Payer category	text	varchar	4	See Lookup Table LT paycat
14	pos	Place of service code	text	varchar	2	See Lookup table LT pos
15	urban	Urban indicator	numeric	int	1	1 if Oregon urban ZIP; 0 if not Oregon urban ZIP; blank if missing
16	dx1	Principal diagnosis	text	varchar	5	See current ICD-9 documentation from CMS
17	px1	Principal inpatient procedure	text	varchar	4	See current ICD-9 documentation from CMS
18	proccode	CPT or HCPCS procedure code	text	varchar	5	See CMS web site for HCPCS codes; CPT codes are proprietary
19	status	Discharge status	numeric	int	2	See Lookup Table LT dstat
20	los	Length of stay	numeric	int	5	
21	ecode	E-code	text	varchar	5	First occurrence of ICD9 diagnosis code that begins with "E"
22	hcg	HCG code	text	varchar	4	See Lookup Table LT hcg
23	msdrg	MS-DRG	text	varchar	3	See current MS-DRG documentation from CMS
24	Citation: Oregon Public Use File. Available at: http://www.oregon.gov/oha/OHPR/RSCH/docs/All_Payer_all_Claims/PublicUseDataLayout.xls . Accessed on July 2, 2015.					

4. *Build validation into the data release process.* The APCD’s data governance structure should clearly define data validation processes before releasing APCD data in a report, dashboard, or other output. When developing an APCD, states should decide who will be responsible for managing data validation during the release process. Some states (Minnesota and New Hampshire) engage an external vendor to handle all data validation, while other APCD administrators perform this function in-house. States should also develop a plan to address issues identified during the data validation process, “message” these issues appropriately, and manage the resolution.

A critical stage of the data validation process is ensuring the data outputs match the data source. States that release state agency data from the APCD should allow the source agency to review and approve the data prior to its release. This “preview” allows the state agency to validate the APCD analyses against the agency’s own dataset and flag any inconsistencies. Similarly, the APCD team should share APCD

reports that measure health care providers' cost or quality of care with those providers. For example, Colorado's APCD team gives hospitals thirty days' notice to review and validate the APCD's hospital price data before releasing the information on the public-facing website. During this time, the APCD team also meets with the hospitals as requested to review and discuss any issues or concerns that are raised.¹ The APCD data manager should coordinate these "previews" of forthcoming reports with providers and/or state agency data sources, and respond to comments or concerns.

As an additional validation measure, some states appoint a scientific review committee specifically to validate academic research that uses APCD data and ensure that the publications accurately interpret and represent the findings. Arkansas and Washington have built this type of committee into their APCD legislation, although they have not yet implemented it.

Applying these data quality and validation processes at all stages of the APCD will ensure the quality and integrity of data within the data mart, as well as the accuracy and reliability of all data outputs.

¹ *Colorado Medical Price Compare: Overview and Frequently Asked Questions*. Colorado Center for Improving Value in Health Care. Available at: http://www.civhc.org/getmedia/55d25b43-c61d-4d68-adcd-4338a7835f78/CO-Med-Price-Compare-Overview-and-FAQs_Final.pdf.aspx/. Accessed on July 2, 2015.