DATA QUALITY WORKSHOP SUMMARY
Break Out Discussion Summary and Recommendations
October 12, 2018

Data Quality Breakout

The scope of the workshop was directed toward both of the major health care reporting systems: statewide hospital discharge data systems (inpatient, Emergency Department, and Ambulatory Surgery (IP, ED, AS)) and All-Payer Claims Databases (APCDs). The participants represented data officials from state agencies that manage one or both of these types of data reporting systems as well as those managing private data systems and the vendors that support these public and private systems.

What is a definition of data quality? Data quality is an assessment of a data's ability to serve its purpose in a given context. If you apply valid statistical techniques, the user will be able to conduct accurate/correct analysis

Characteristics of Data Quality (adapted from HSRI):

- Accuracy: degree to which fields are close to their true value
- Completeness: degree to which expected information is received/fields populated
- Integrity: degree to which information is valid, consistent, reliable
- Relevance: if the information important to the users
- Timeliness: Data currency after data are cleaned, new data integrated, and extracts developed/released

Overall challenges guiding the scope of data collection and validation policies and practices should be guided by the realities of claims-based data:

- For APCDs, there is the issue of what payers need to adjudicate a claim and what is kept and stored in their payer data warehouses.
- For APCDs, submission to state APCDs is a secondary consideration. Payment is primary purpose of claims.
- National payers benefit from a standard set of data elements and edits across states
- States need to know what vendor checks are applied to be able to have a conversation and educate data users
- Challenge: missing data/exclusions such as VA, IHS, OPM, Part 2, and ERISA/SF data
- How important is it to have flexibility to use home-grown codes? State-defined fields occur, despite data standards (eg X12, NUBC)
- Updates to systems, improvements may cost $$$, who pays for this?

INCOMING/DATA INTAKE: edits that are applied at the time of data submission prior to importing data into the data warehouse

Key themes and related practices from the discussion include:

1. Documentation
2. Standardization

3. Compliance
4. Stakeholder Engagement

1. Documentation: Good documentation promotes consistency and transparency in reporting While there are always exceptions, there needs to be an agreement between the data agency and the data submitters on the need for and scope of common set of data/validations on intake

- Submission manuals in the form of Data Submission Guides provide detailed technical submission specifications.
  - Keep all historical versions of the data
  - States issuing Request for Proposals (RFPs) should incorporate expectations for audits up front
  - It is important to apply a standard set of edits, but there may be a need for some customized edits for local codes/submitter codes to be added to reference code sets (e.g. specialty codes)
  - Annual review of data submission requirements recommended

POST INTAKE/VALIDATION CONSIDERATIONS

- Post processing definitions are needed
- Draw on other data sources to use as reference checks:
  - APCD to hospital discharge
  - Rx utilization
  - vital records
  - cancer registry
  - Provider directories
- Other validation considerations: When reviewing submitted data, assess for:
  - Reasonable
  - Match across files
  - Average age, number of records/claims/members
  - Member match rate
  - Longitudinal analyses
  - Market analysis
  - PMPM change
  - Seasonal variation/quarter to quarter (% threshold) and quarter from previous year

- Value-added enhancements:
  - Addition of groupers (APR-DRGs, MSDRGs, drug classification)
  - Master Patient Index MPI)
- Data transformation (recoding)

COMPLIANCE CONSIDERATIONS

- The will/the authority/political cover to enforce standards. Fortitude: how hard to you want to push to get to 95-100%
- Maybe tightening legislation to fill the gaps, Clear penalties
- Compelling submitters to send “good” clean data because it benefits submitters themselves

- Variation in data across states may be related to data system governance. Mandated reporting requirements provide leverage to enforce requirements while voluntary submission relies on submitters to submit correctly and other incentives for improvement in reporting are needed.
- Develop tool to make it easier to comply (comparative reports, feedback, portals)
- Provide reports back to hospitals/payers, metrics (health care)
- Carrier/provider reports benchmarked against peers, may be "post-production": informs the incoming validations
- Based on payer-specific history (post-production) establishing some payer-specific thresholds
- Codes/standards are there, there is a political dance/dynamic for how far do you go, how hard to you push?
- Fines? Are fines compelling?
- Encourage submitters to use the data that was submitted
- Connect the submitter within organization to user/audience within the organization

STAKEHOLDER (SUBMITTER AND USER FEEDBACK) CONSIDERATIONS

There was general agreement that submitters and users are critical stakeholders and it is essential for the agency responsible for the data system to actively reach out to both submitters and users in all aspects of the data cycle (planning, intake, improvement, usage) and establish ongoing mechanisms for doing so.

Submitter Practices Identified

- Communication with Submitters. Submitters provide explanations of 'anomalies' frequency
- Standard reports to submitters
- Increase data use by submitters
- Fixed schedule, clear timelines
- 2 weeks internal review
- 30 days submitter feedback
- 30-45 days for processing and follow up/shorter timelines
- Strong feedback loop with submitters

User Components Identified

- Data User Feedback
- Provider analysis
- Feedback from internal data users
- Power users, regular users
- Provide more detailed analysis
- Opioid use researchers helped raise data redaction issues in 42 CFR records
- Created data user specific validations
- Biannual meeting with data users
- Analyst cafes to get feedback
- Quality sharing with Data users
- Lunch 'n Learns: share research and data concerns/limitations

- Standardized variables in public use files for public use files, different use cases
- Standardized extract variables by type (hospital/APCD/etc)
- Standard data products

BREAK GROUP on Post Data Validation (Notes from HSRI)

1. What is actionable?
2. How do we make improvements?
3. What is post vs pre validation?
4. Post validation
5. What is an analytic dataset
6. Business rules
7. Validations post processing
8. Sharing what people should be or would like to do

Describing Post Validation and Potential Checks

- Maryland: After pre-validation, then run data through groupers, MS DRG, APR DRG, APC, PPC, drug classification then value added variables (race etc.) then master patient index for every discharge and visit, then it comes back to us for analytics
- Freedman - these would be called data enhancement that would come after post processing. Reasonableness checks - do metrics match across file submissions, is this calculation falling between range, hundreds of reasonableness checks - flag if linkages - across files like eligibility and claims/member match rate are above cut off. Compare to metrics to previous quarter and same quarter in previous year's submission. Also these checks include metrics such as PMPM.
- Virginia: called them longitudinal edits

What are the checks?

Established cutpoint. Some groups used z scores and standard deviation.

- Freedman: Quarter to quarter - payer - twice a year, and same quarter in previous year - Freedman
- Florida: 1000 facilities, 330 hospitals, 600 AD free standing ambulatory. Process all data ourselves xml format. 700 edits for outpatient, run through auditor software, vendor calculates a norm report, elements selected for this report, average and standard deviation for a year worth of data. Then a threshold report - more of an alert that this may be a problem, take it down to ground zero, certification based on value - distribution for each of the elements. When going back to facility, ask them to validate what is seen on the norm report. For example race - detected a lot of mapping errors, hawaiian added. Lots of hospitals changing vendors - not all hospitals do the same thing; even with same vendor there is variation, depending on the analyst at that facility

What do you communicate back to your submitters? What is actionable?

- **Freedman - Oregon - Milliman**
    - Level II and III check
    - Level II is done semiannually, counts everything, standard deviation - measure by measure is different, might be proprietary to Milliman

    - Spreadsheet is summarized to point out areas, hound them to reply
    - Flag more things than is added to summary spreadsheet
    - Level III - enrollments, PMPM, utilization etc. - ask submitters to validate it - leave it up to them to respond "provide this data to you as a courtesy, have 60 days to reply back"
- **Maryland**: since data is used for payment, monthly reports for metrics relevant for payment, market shift type of analysis, service areas moving from one hospital to another; they have 3 months of the data to resolve
- **Vizient**: strong incentive to correct those data because used for performance groupers. Hospital - caring to be just under the threshold
    - Payer: Sometimes allow for delay in replying, otherwise then a penalty is applied
    - Having metrics that align with our data users
    - Validation - age, race ethnicity - discrepancy between hospital data and birth records - requires a linkage between data file
- **Freedman**: Just started this in Oregon - call this Level V - compared across data sources - against hospital discharge, prescription drug utilization records, cancer registry - identify
    - Identify all elements present in all these sources, identify denominator ,should be similar
    - The goal became to understand why they don't tie out
- **Kathy Hynes**: APCD to understand who payed for something, hospitals for race ethnicity
    - Clinical - hospitals better since APCD is using info just for payment
- **Maryland:** compared the data to Medicare - CMS to hospital discharge data; some hospitals are ok totally depends on how hospitals are identifying payers
- One source of validation not used by Vizient is registry data. Same thing when compared to cancer registry - for example end of treatment is different - crossed out metrics that did not match between the two sources

Time between when data intake and post processing validations

- 3 months - Arkansas
- 45 days - Maryland
- 2 week internal review, then give the submitters 30 days - Virginia; different process for discharge data - 2 months from receipt, until edits, send edits back, all adjustments are made then released
- You can't give payers less than 30 days
- Annual files - 120 days for review previously, now improved to 45 days
- Industry standard for data processing, and requirements for vendors - has decreased in past years

<u>Lessons learned for streamlining/fines:</u>

- Maryland: monthly reporting back to hospitals; several steps including third party vendor, got intro sprint with them, next 2 weeks must be done, fixed schedule - nailed this down
- 30 days after submitting get a report back
  Give payer 2 months to respond, during that time the payer can submit and resubmit, time to give feedback, but encourage payers to respond sooner, relationship building with the submitter, Imposed penalties to one payer - people got fired because of that - give them 6 months; $1000 per day per file delayed
- Florida: fine based on how many times have you been delinquent prior to that;

- Maryland: used to be $500 fine; started to include this in global budgets so that facilities are aware of consequences of not submitting

Integration of data user feedback?

- Virginia: Feedback about providers
- Maryland: Feedback from the health department that is using the analytic file; example unknown race
- Florida: power users, fine tooth comb, finer payer categories used for their research; or for example auditing number of procedures (one hospital with decrease in number of colonoscopies)
- Freedman: many data submitters are incorrectly interpreting 42 CFR part 2 substance use disorder treatment - but also opioid prescriptions now developed validations to make sure that the data should be there is there very recent, can't speak to whether effective
- Analyst cafes once a month, listservs - to get feedback
- Lunch and learns in Rhode Island - research, data concerns, informed data validations, if data cannot be improved then add to the data user guide

Communication with Data Users

- Florida: developed a user guide and 3-4 how-to videos especially for ASC, low tech

How to make the call to not release data

- Catching internal processing errors
- Virginia: quantifying the issues, decide whether to take it down
- Rhode Island: offered refreshes for free, given refunds, for public reports we have suppressed data

Wish list

- Arkansas: more discussion among states about carriers, like United and see if we have common issues
- Maryland: standardized variables that are contained in public use files; give out a lot of data - non PII, so it would be good to find common elements and standard validations
- Standard data products, limited data sets, public use files - but difficult to certify this; specs for that are publicly available - Washington is also making that available
- Social Security N
- may want to consider the last 4 of SSN may be very useful in identifying people

Virginia: has a log of the elements that people look at the most in their system
a few fields that get used all the time, then rest only 3% of the time

Goal: Develop consensus and guideline best practices for all states, all payers, with some language around exceptions.

COMMUNICATING DATA QUALITY TO USERS

What do folks need to know?

- Data submission guide
- Data dictionary for available fields
- Data book with basic information (#claims, #submitters, #members)
- At time of release and notification if corrections/changes have been made and new files issued
- Missing data (months, carriers)
- Identify the laws/policies as dictate release so users understand reasons
- *Data discovery log, helpful before the request
- * ”Snapshot of data as of today”—disclaimer
- Vendor QC report: basic statistics, running list of subscribers
- For identified data issues, define the scope of problem(s) and suggest work-arounds and fixes
- Release notes about the data, including versioning information, at the time extract was generated
- User manuals for limited use files
- Submitter-defined code cross-walk in the form of excel file and/or separate look-up file/table
- Data User Groups
- Agency presentations/guided learning materials
- Push and Pull communications:
    - Pull: Systematic way for users to submit questions. How and who to contact.
    - Pull: Direct email and/or a service desk to triage/vet questions
    - Push: follow-up communication from User Group meetings posted on website and email
    - Push: You tube and Data Academies—classes to potential users of APCD
    - Push: Data Users Groups and quarterly updates
- Power/sophisticated users may require close personal communication
- SAS code developed by the agency---calculates member months for eligibility file for users
- Agency internal training for those working on and with data set
- Include payer history in submissions

What do people need to know about the data?

- Difficult decisions about how much and what to communicate to users and when
- A state posts a data book with detailed information about the files (total submitter, total claims, etc.) this information helps users interested in requesting data and it guides the release process

How often do you communicate/cite the law?

- A data cover sheet includes data discovery log, documentation of data quality issues, description of the issue and extensiveness of the issue (payers/claims affected), options for work arounds/status
- Quarterly updates provided to people with extracts are provided
- Disclaimer around data status (what submitters sent at the time, it's what you get)

When is a re-release indicated?

- If it's the agency's "fault", a rerun/repull will occur. If it's a client's change of mind, it's a conversation with that client. If it's a submitter issue, it depends---many variables to consider
- If there is an agency mistake in pulling data, a new file should be issues
- If the requestor changes his/her mind, no new file need be issued unless there is willingness to pay
- If submitter mistake: it depends. Small issues may just need notice to data receivers. Big issues may warrant re-issue.

NAHDO DATA QUALITY FORUM NEXT STEPS:

Standards

Call to action:

1. Business case for CDL/PACDR (many business needs can be met through these standards)
2. Frequency or percent filled fields for existing states
    a. CDL needs work before wide use
    b. Front matter
    c. References

Own CDL governance, support PACDR. The CDL and PACDR send a message to hospitals and payers that this is the expectation. States aligning can make it easier for national submitters.

- Starting point:
    - CDL with national standards for codes and PACDR (transactional)
    - Then percent threshold---later

Cross-state Metrics to Compare Data Quality---"So we all are not working in a vacuum"

- Share common feedback reports and common issues with national payers

EXAMPLE CROSS-STATE REPORTS:

	NH	MA	HI
Patient last name	95%	100%	70%

- Comparison across states PMPM, other measures state-to-state
- Cut points based on distribution (Z scores)
- Best Practices: this is what states have/can achieve, then new states can figure out how to get there

- For existing states/new states, may want to share current thresholds—sharing how states are performing and proposing goals
- Common denominator: get information about possible thresholds for data submitters, get them involved again

National User Group: Participants recommended a NUG modeled after state user groups

- National Users Group: submitters and frequent users of HDDB, APCD and groupers
- A group of states should work to develop 'best practice' educational materials for data users and to then present these to other national groups (AcademyHealth, NASHP, etc.)

NAHDO WORKGROUPS

- Standards Workgroup is needed, including payer participation
- NAHDO members IT group for IT sharing: data base, data wonks, tools, auditing states and analytics for developers

NEXT STEPS: BREAK-OUT SUMMARIES

Data Intake

1. Renew discussion on use of CDL

2. Develop standard set of top (10 or so) data validations to compare across states, collect the results each state receives for each metric

3. Develop data quality sub-committees/working groups, could include- CDL, IT Infrastructure, Analytics and Submission

Post Intake

1. Compile information across states on timeframes for data validation and release

2. Compile information across states on fines assessed for data quality (I actually have some data on this one already)

3. Compile information across states for which metrics are used to provide feedback to data suppliers

Communication

1. More standardization and clarity are needed in data dictionaries especially in how they relate to data submission guides

2. There is great interest in developing standardized trainings for working with APCD/claims data that can be used across states

3. Compile best practices, topics can include- release notes template, data issue log template, data disclaimers