Notes from Calls to other States about the Online Scorecard

Introduction

As Connecticut is beginning the creation of it first online health care scorecard we have had a series of discussions with other states who already publish scorecards. Calls took place with Minnesota (7/27/16 and 8/8/16), Washington (8/1/16), Wisconsin (8/11/16), California (8/29/16), CA UCSF (9/27/16), CA Healthcare Compare (10/11/16) and Maine (11/4/16.)   Attendees included representatives from the State being questioned, CT SIM PMO, UCH Evaluation Team, ONC and CMMI. CT submitted a set of questions about procedures prior to the call. These questions formed the framework of the calls.

Summary

During discussions with these states a few commonalities were evident:

- Relationship building. States spend a great amount of time and effort in a "buy in" period to receive input from and gain the trust of providers and provider organizations. They all stated that this time was worth the effort as they do not get complaints about scores.
- Users varied. Users of the scorecard varied. Primary users were not healthcare consumers.
- Data validation. Data validation, including audits, was a regular part of data collection cycles.
- Attribution. Attribution methods varied. Often patients were attributed to the provider that they saw most frequently. Alternatively, sometimes provider organizations assigned patients to providers when submitting data.
- Risk adjustment. Most States did not perform risk adjustment – except when required by the measure. The one state that performed risk adjustment established a risk adjustment committee to guide the process and also presented both adjusted and non-adjusted scores.
- Staffing. Staffing for the scorecard project was considerable in each state (between 3 full to 20 partial FTEs plus vendors.)   Staffing needs were higher for initial states and reduced over time.
- Vendors. All States used vendors for the creation of the websites and underlying data repositories.
- Impact assessment. Few states performed impact assessments other than using Google Analytics and informal discussions. Maine also used a user survey posted on their site.

Decision Points

The following decision points need to be discussed to guide the development of CT's development online health care scorecard:

- Purpose and users of scorecard
- Attribution method
- Risk adjustment
- Scoring
- Presentation

Detailed Notes

Purpose:

MN: Quality reporting and measurement system- some private sector and some public sector. Effort began with State level health reform in 2008.

WI: To allow for comparison among providers.

WA: Source of debate. Scorecard is for purchasers and consumers to make better informed decisions, to allow providers to see their performance and drive quality improvement and for health plans use data to compare their results to results of state.   Began data aggregation in 2007 with voluntary data submission. State passed legislation in 2014 for public reporting on a common measure set and started reporting measures in 2015.

CA Px Adv.: Hospital reporting data has been around for decades, while the open data portal is a few years old. The focus is on transparency and on allowing consumers to make informed health decisions.

CA UCSF: Driving health care quality. Originally the scorecard was envisioned for healthcare consumers, but few consumers have utilized it.

CA Ins. Dept.: Began from the Federal Rate Review grant and the desire to add quality data.

ME: Originated in response to mandate to take APCD data and produce consumable information on health care costs. Primarily intended to be used by consumers making health care choices.

Scorecard Development Activities:

MN: Multiple advisory committees and user testing. Extensive discussion during development. Measures are updated yearly and are extensively reviewed in committee and a public comment period.

WI: Early on this endeavor was considered risky in WI. They went very slowly and engaged physician leaders though a measurement advisory committee to create a plan that was sustainable. They went through an extensive Beta testing of the process, including the data collection, analysis and denominator/attribution methods.

WI believes that their success today is largely due to the fact that providers submit their own data and can validate results themselves. Groups with low scores do not blame the methodology.

WA: Held many meetings, received private comment from and exchanged emails with stakeholders. Convened committees and subcommittees including a Board of Quality Improvement made up of leaders in hospital and physician community. Decided to use only NCQA measures with national benchmarks. Seven years elapsed between the initial

2

voluntary effort to begin reporting and the formalized legislation that created a common measure set and public reporting.

Also spent a lot of time preparing and researching (websites, literature, etc.) about possible scoring methods.

CA Px Adv.: Inpatient hospital reporting data has been around for decades and has already been extensively used by the hospitals. The open data portal was supported by 12 agencies working together. They have a standing technical advisory group with members in the research and academic community. Most questions about method get hashed out there.

CA UCSF: Scorecard stakeholder group and its advisory committee spent many meetings reviewing plans, making decisions on functionality and presentation. They also held user testing via focus groups and 1:1 sessions observing users interact with the site and asking questions. User testing was conducted to refine description of measures and other text on the site to improve accessibility.

ME: Received federal grant to enhance public reporting. Convened multiple workgroups including a consumer advisory group from which they elicited a great deal of feedback concerning measures, scoring and presentation.

Scope:

MN: Health outcomes and care experience. Target audience is mainly purchasers and providers.

WI: 40 ambulatory centric, 6 patient experience, 3 surgical, 3 cardiology measures. Report on 60-65% of practices in WI. Main gap is in small practices.

WA: Currently 52 measures covering prevention, acute and chronic care.   Common measure set is not static but is updated as interest, science and data sources become available.   Report at state, county, health plan, hospital, and clinic level.

CA Px Adv.: Open Data Portal includes all publically available data from 12 CA Health and Human Services agencies. Inpatient hospital reporting includes diagnoses and procedures and, for the past decade, outcomes for coronary artery surgery, hip fracture and several other reports using ARC methods. For the past 6 years they report at the surgeon level. Reports are published via .pdf on their website.

CA UCSF: Beginning in 2002-3 as a site rating nursing home quality of care. 2-3 years later they added quality ratings for hospitals. Hospital ratings include patient experience, safety and overall re-hospitalization rates as well as ratings on specific conditions and types of care.

CA Ins. Dept.: Hospitals and large medical groups. Provide data on 7 healthcare domains- 3 hospital and 4 medical group domains. Under each domain are overall cost, quality, patient experience ratings. Drill down provides measures under each domain.

ME: Health care costs and quality for hospitals, labs, and clinics.

3

Data sources:

MN: Medical records from practices (EHR) and CAHPS

WI: Data comes from physician practices and consists of both EHR and administrative data.

WA: Claims, survey and registry data

CA Px Adv.: Open Data Portal uses publically available HHS data.   Inpatient procedure outcome reports use a combination of a hospital discharge database and clinical data from the hospitals and death records.

CA UCSF: All publically available data sets.

CA Ins. Dept.: No raw data utilized. Data comes calculated from sources such as CMMI, CDC, etc.

ME: Cost data comes from claims (APCD) and quality data comes from public sets- CMS.

Data collection/validation processes:

MN: Providers submit data in three waves during the year via a secure portal. Detailed data submission guide is used. CM audits and validates it via remote access of practice data. Providers have 2 weeks to review results before they are published. Use an audit and appeal committee.

WI: Submissions come directly from physician offices which upload data into a clinical data repository. WI provides practices with detailed specifications for the submission. Practices give WI the query language used to pull the data along with the data. The Clinical information manager validates the code. Field audits are performed if they receive partial data sets, for some measures every period, or if they see a large increase or decrease in performance for a practice.

The measurement period ends December 31 each year and practices can begin uploading data any time after that. Providers are given a preview of results one week prior to publication. They can login to the repository and run their own analyses.

WA: Two step validation process. The first step is done by their data vendor who does some checks on missing values, tallies equal what is expected, etc. Second, health plans review aggregated claims data. Providers are given 3-4 weeks to review data/scores (?). Time spent validating data is long, but is an important step in achieving provider faith in scores.

CA Px Adv.: Inpatient data and death records are received as flat files every 6 months. Hospitals package data and submit to a legacy Oracle system. An extensive process of staff driven and automated validation takes place next. The data is of high quality as this process has been ongoing for a very long time. They used to do chart reviews but have found that not cost effective.

4

CA UCSF: At the beginning they had a process for some measures, i.e. ICU mortality, that involved direct submission from hospitals (nurse chart reviews) and was followed by report validation and audit. This was very time consuming for the hospitals so they switched to measures that were based on publically available data and do not do data validation.

CA Ins. Dept.: They do not use any raw so only perform "desk audit" data validation- i.e. looking at outliers, patterns of missing data, etc. CA does ask the source organizations about questionable data points. Any data points that are confirmed as errors may be suppressed.

ME: Providers input data into associated APCD. A set of validations are done automatically at the time of submissions and a second set are done following. Data validation is done at both the structural and data element level.

Attribution method:

MN: Patients are attributed to the providers who patients have seen the most.

WI: 1:1:1 Relationship of patients to providers to practices. Two methods for attribution are used. The provider group can attribute patients to providers. If this is not done then WI does the attribution. In the latter case for a patient to be attributed to a physician they need to have had at least two visits to that provider in the previous two years, with at least one in the measurement year. Data is siloed so patients cannot be matched across providers, so it is possible that a single patient may be in more than one provider's denominator.

WA: Patients are attributed to PCPs, teams, and prescribing providers. They attribute patients to the PCP who has seen the patient the most times. Teams include all providers who have seen the patient for a chronic condition. WA reports that these methods are used fairly regularly across the country. They have rosters that link physicians to clinics and medical groups.

CA Px Adv.: Not applicable. Data comes with patients already associated with physicians and hospitals.

CA UCSF: Not applicable.

CA Ins. Dept.: Not applicable, no raw data utilized

ME: Not yet communicated- will be provided via follow up documents

Scoring:

MN: Present performance as actual and expected against state average.

WI: Original goal was to benchmark providers against each other. They had no consideration of

rollup of scores. Additionally, since they use home grown measures with denominators that do not match up to national benchmark denominators, they cannot easily find national benchmarks against which to rate providers.

WA: Use only NCQA measures with national benchmarks. Report performance rates with confidence intervals.

CA Px Adv.:   Hospitals and surgeons are rated as below avg., avg. and above avg. relative to others in
the state

CA UCSF: Initial thoughts to rate hospitals against each other were met with resistance as there would always be hospitals in the lowest quartiles regardless of whether their care was adequate. So, they adopted a system of comparing performance against state or national benchmarks. If measures fail to show variability or space for improvement then they are considered "topped out" and retired for new measures.

CA Ins. Dept.: Use a 3 or 5 scale rating system to translate performance into an "understandable" rating. When choosing a rating system they stressed the importance of gaining provider buy in on the rating system over trying to find a single "correct" system.

ME: Present median costs for rated entity and provide ME average as comparison. Quality

performance is expressed in cell phone bars and is relative to national and state benchmarks.

Risk adjustment:

MN: Have risk adjustment committee. Present actual and expected ratings.

WI: No risk adjustment is done other than what is required by measurement specifications. They would do this if they were reporting at the level of the individual providers. Part of the reason is historical- the way the data was originally collected they were unable to do any risk adjustment. Currently they have no complaints about the lack of risk adjustment. They do actively seek to understand the differences in care to WI subgroups.

WA: No risk adjustment other than what measures require.

CA Px Adv.: Risk adjustment is done but details were not available during the call.

CA UCSF: They do not add any risk adjustment but utilize risk adjustment that is built in to the data/measures.

CA Ins. Dept.: No raw data utilized so no risk adjustment is done beyond what source organizations do. They sometimes do stratify results, i.e. group results for larger or urban hospitals. Stratification is in response to provider requests to be compared only with like groups.

6

ME: No formal risk adjustment but perform "grouping" which categorizes claims into different levels of severity. More
details will be sent via follow up information. Grouping has eliminated the effect of patient sickness level on most measures.

Frequency of update:

MN: EHR based measures are updated yearly, but care experience every other year.

WI: Twice yearly in May and December

WA: Yearly

CA Px Adv.: Yearly

CA UCSF: Did not ask.

CA Ins. Dept.: Yearly, although some measures have required quarterly updates.

ME: Cost is twice yearly, quality yearly

Software:

MN: Wowza.

WI: Ancilla Partners built and maintain data repository and website. Unsure of software used- built from Ruby . Software custom designed.

WA: Orchard, Tableau, Umbraco for content management.

CA Px Adv.: Oracle for data collection, html for website- moving towards wordpress. No online interactive scorecard.

CA UCSF: Digitaria for website which is custom built in HTML5 and Javascript. They use the Google Maps API for geography based searches. Content management uses SiteCore

CA Ins. Dept.: Used what the vendor recommended- MySQL, Java, Apache, etc. Vendor is an external company called Ringful Health.

ME: Not yet communicated- will be provided via follow up documents

Support needs and staffing:

MN: Eleven full FTEs for organization. Programming done in house. Staff includes:
2 programmers

Few analysts
Communications specialist
Data architect

WI: Budget of 1.6M, covers this project and a second one. Scorecard uses 3 full time staff plus a vendor who created and maintains the data repository and website

WA: Twenty full FTE staff for entire organization plus uses a vendor to help with software development. Project requires at least .5 FTE for consumer website and media questions.

CA Px Adv.: About 25 staff involved in data collection and regular communication with 450 CA hospitals. Small unit of research staff including a director, 2-3 PHD level researchers and 3-4 research assistant/analysts. They also partner with U of CA to bring MD and PHD researchers in the develop studies.

Most questions and complaints are regarding timeliness of data (18-24 months old.) Providers with low scores will often respond that the data is too out of date and that they have since made improvements.

CA UCSF: 2-3 full staff plus vendors. Total costs run $600-700k per year, with additional costs during startup phase.

CA Ins. Dept.: Depends on status of project- initial 6 months took 4 full FTEs at Consumer Reports, currently have 4 staff with largest FTS from a project manager (10%) and data analyst (20-25%). Will use more staff time during periods where measures are being added or other changes are being made. UCSF also employs a few staff on project.

ME: Staffing requirements were much greater during the initial launch of the scorecard and have dropped over time. More details will be sent via follow up information.

User testing/impact assessment:

MN: Did not ask

WI: Google Analytics

WA: Google Analytics. Users are Providers, care mangers, medical groups, consumers, and Public policy officials

CA Px Adv.: Did not ask

CA UCSF: Impact assessments have not been done in the last 2-3 years. Previously they looked at the web traffic to see where users originated from, what pages they visited and for how long to investigate the impact of advertisements. They found that users brought in via advertising did not stay on the site long. Those that found the site through google searches utilized the site in a more substantive way.

CA Ins. Dept.: Track email open and click through rates and also track web traffic. Do not

collect information on specific users. Previous research has demonstrated that these sites are used by patients to check on their providers, but also for patient doctor/hospital choice and by providers for referrals. They do no direct impact assessment, but have had anecdotal reports that their site is used by providers to track their own performance and drives improvement efforts.

Consumer Reports has a lot of knowledge about how to set up the website in a user friendly way, but they additionally did cognitive testing with different types of users (a person seeking a knee replacement, a person looking for a hospital for childbirth, etc.) to see how they would use the site.

ME: Google analytics tracks unique users. They have a survey on the site to collect information from users. They also receive emails and some phone calls with questions and concerns. Going forward they would also like to publish a report every 2-3 years detailing the changes in health care cost and quality in ME.