Health Plan Quality Measure Inventory and Methodology

(December 10, 2012 meeting topic)

Dann Chapman: University of Minnesota

I think this whole measurement thing is fraught with peril. It's tempting to say 'measure everything on this list', but; how would it get weighted and aggregated to be of value? And, I expect, that approach would be horrendously expensive.

Whoever mentioned measures not being clinically useful made an extremely important point. This is something MNCM wrestles with regularly, and there's a tension between providers, who sometimes would rather not be measured on things that are hard for them to control, and the value of reporting on exactly those things to push them to try harder. Then there's the fact that few of us on this work group have any expertise at this level.

Having said that, it seems to me that we should focus first on those measures - such as access - that consumers care most about. And that we should focus on those measures that make sense at the QHP level. Quality of care is, of course, critically important. But quality of care at the plan level is virtually meaningless unless the 'plan' only includes one ACO. Anything bigger than that, and even at that level, quality is going to vary considerably provider to provider. Does it help to know that at ACO 'X' they exceed at all these systematic measures? Yes, because it gives me a greater chance of picking a really good doc, who is backed by a system that enables her to provide top notch care. I could, though, still pick the worst doc in the ACO. Would he be better than the best doc in an ACO that showed up lower on the measures? Not necessarily.

I think what I'm getting at is that we should be careful of over-promising what our measures can deliver; either to ourselves or to the public.

Let's focus on those consumer-favored measures. Make them easy to understand. Eliminate any measures of questionable value for picking a plan (might need some expert guidance here). I favor using only measures that are already being used and reported, which also have the approval of one or more of the agencies we've referenced. And I'd like to see at least two suggested configurations of this, built by people who have more expertise than I have.
Sue Knudson: HealthPartners

In response to the last health plan quality measurement exchange committee meeting, here is some additional feedback from HealthPartners. We recommend a set on principles from which to evaluate the measures and further suggest some measures for use. This is consistent with previous feedback at meetings but we thought it would be useful for you to have it in writing.

Thanks,
Sue

Principles:

• Measures should be readily available & already publicly available
• Measures should reflect/demonstrate a plan’s breadth & scope of capabilities
• Measures should reflect a plan’s ability to impact (...either directly or by partnering with providers)
• Measures should be meaningful to consumer
• Measures should not be imputed/inferred
• Measurement & approach should result in accurate, fair, and meaningful comparisons for consumers (goal is not to go to the “lowest common denominator” to ensure there is a result for all or most)
• Measurement should consider regional market dynamics
• NQF measures should be a key consideration for measure selection but exceptions should also be considered (e.g. may often lag most current evidence-base, may not reflect regional definitions, may not address measurement gap areas, etc)

Suggested Measures:

• NCQA Star ratings & supporting HEDIS measures
• Consumer Reports rankings
• HMO/PPO breakouts not helpful for MN market
• Recommend overall rating as well as key quality/patient experience domains to be displayed
• New plan entrants or small plans lacking data need to be clearly reflected as such – “n/a” for overall rating but could list domains where quality exists
• NCQA measures highlight key quality processes/outcomes, patient experience. Rankings bring in the measures along with the accreditation ratings (...which further highlights plan performance).
• Good idea to round out/gaps using Evalu8 metrics but these are not readily available.
• Somehow network adequacy needs to be highlighted. Not sure CAHPs covers it (....access to specialties, appointment wait times, distance traveled, etc). Would help to highlight concerns around “super narrow” networks that present very cost effective but has numerous member “surprises” (e.g., no allergists in network, or access to primary care 4-month wait)
Options for Quality Reporting for 2013 Open Enrollment
(January 14, 2013 meeting topic)

Dannette Coleman: Medica

First, I want to say that I thought Checkbook did an excellent job of laying out the options as well as identifying the pros and cons. I have had a chance to talk internally to our team that is accountable for our NCQA accreditation as well as our broader quality initiatives and they provided their expert opinion which I am forwarding on to you. Please let me know if you have any follow-up questions.

The rankings list is based on statistically insignificant differences among the health plans listed. The difference between a plan ranked 50th and one ranked 60th could be only a fraction of a point. As a result we do not consider the rankings list to be an accurate reflection of performance.

CAHPS is not a particularly sensitive instrument. It also measures a number of variables that health plans can’t control such as satisfaction with provider behavior.

Of the three options, the scorecard would be our preference. Although, not ideal. It displays the most accurate reflection of plan performance as it incorporated HEDIS as well as CAHPS.
Jim Naessens, ScD

On the initial plan comparison summary screen, the Exchange presents for each available plan an overall quality measure.

This overall quality measure is the “Rating of Health Plan” score from the CAHPS survey and has the heading “How Members Rated the Plan” or similar.

As soon as possible, the Exchange lets all carriers that are likely to want to participate in the Exchange know that CAHPS scores will be used in quality reporting.

If some of the plans cannot produce CAHPS scores for commercial enrollees but can produce such scores for Medicaid enrollees, the scores for the Medicaid-only plans are shown along with the scores for the commercial plans with a reasonably prominent note about data being reported on two different populations.

However, since we expect differences between Medicaid and commercial plans (and both to differ from the exchange group) I propose that CAHPS scores be reported separately for Medicaid and commercial providers by carrier. Where the carrier has no measured experience, the report should reflect that the data is not available. In this way a consumer could see whether the carrier performance varies across population and could compare more carriers across consistent populations. Something like the following, as example:

**HEALTH PLAN RATING**

<table>
<thead>
<tr>
<th>Plan Type</th>
<th>Carrier A</th>
<th>Carrier B</th>
<th>Carrier C</th>
<th>Carrier D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Commercial, Medicaid</td>
<td>Commercial, Medicaid</td>
<td>Commercial, Medicaid</td>
<td>Commercial, Medicaid</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>4, 3</td>
<td>3, 1</td>
<td>2, 5</td>
<td>NA, 4</td>
</tr>
</tbody>
</table>
I want to follow up on the conversation that took place at the Measurement and Reporting Work Group on Monday this week.

I would like to reiterate my opinion that displaying the fairly detailed quality metrics outlined in the discussion document only for plans with an existing commercial base is not appropriate. I come at this from two different rationales: 1) The existing commercial market is a very different population overall when compared to what we expect to be participating in the exchange marketplace (especially the individual market). The current individual marketplace is a relatively healthy (due to underwriting) population who are likely self employed and financially stable. We expect the exchange individual marked to be a mix of the existing market with a large infusion of currently uninsured (either due to affordability issues, health status issues, or just young and invincible) as well as people who are currently in existing public programs and MCHA. 2) Displaying quality information for some plans on the exchange and not for others potentially creates a less competitive marketplace. I am concerned that it would imply to potential consumers that plans with data displayed are higher quality when all it really means is that they currently participate in the commercial marketplace.

As a potential alternative for 2013-2015, the exchange could consider a link to the health plan web site or other shared space that would allow plans to post their own quality information as they choose which could include Medicare data, Medicaid data, or commercial data.

Thank you for your consideration.
Allie Coronis: Allina Health System

Thank you for allowing us to provide comment on the options related to public reporting that were provided by Consumer checkbook. I was most interested in different aspects of Options 3 and 4. I appreciate Option 3, Reporting NCQA's Accreditation report card, and like the information display in Figure two, although I think it is a disadvantage that consumers cannot drill down to see the measures contributing to the star ratings. I was also interested in some aspects of Option 4, particularly the transparency and ability for drill down.

After reviewing this document again, and the notes I took during the meeting, I would be most supportive of Option 5, using a combination, as it will allow for more flexibility and options.

I think I share the same opinion as a number of other workgroup members regarding Option 1, in that it would be undesirable not to have any quality data, where the consumer chooses solely on cost.
Phillip Cryan: SEIU Healthcare Minnesota

Thank you for the opportunity to offer feedback on the memo Consumers' Checkbook put together on options related to public reporting of quality data for QHPs for the Oct. 2013 open enrollment period. The memo was very helpful, and we found the “advantages” and “disadvantages” sections especially useful for clarifying the trade-offs of each option. We also appreciated that the memo underscores the importance of offering quality data not only to promote informed consumer choice for its own merits but also because informed consumer choice can drive continued improvement in the health care system. We don't have a firm recommendation in favor of a particular option, but would like to offer the following comments.

**Option 1 should not be under consideration.** Consumers must be presented with quality information during open enrollment 2013 to help them compare plans on criteria other than price. One of the primary ways the Exchange will offer value to consumers and small businesses is by helping them identify high-value health plans and make more knowledgeable decisions about their choice of coverage. An indispensable element of identifying high-value plans and making good choices is access to clearly presented information on plan quality. The 2013 open enrollment period is the crucial time when much of the population eligible to enroll through the Exchange will form their opinion of how useful a vehicle it is for helping them make meaningful comparisons among health plans. Though none of the options for presenting quality data in October 2013 is perfect, all are preferable to having no quality information presented in the opening year of the Exchange.

**Using only the CAHPS “Rating of Health Plan” question wouldn't be valuable.** The memo points out that this single measurement would present variation that would be interesting to consumers, but we believe it would be at best an inadequate rating and at worst misleading. The problem is not only that plan price influences people’s responses to this question (possibly resulting in a misleading correlation between low-priced plans and a high rating), but also that there's no “drill down” to provide information about the dimensions of quality an individual is interested in. Consumers who expect the Exchange to help them make a choice based on plan quality characteristics they care about and that are pertinent to their particular health needs would be disappointed by this meager offering. Also, using a single reporting metric wouldn't maximize the opportunity to use the Quality Rating System to promote continued improvement in the health care system.

**NCQA's Accreditation Report Card is easy to take in at a glance and facilitates clear comparisons among carriers.** Keeping in mind that research shows many consumers won’t choose to read detailed information about quality, this approach seems to provide about the right balance of detail and simplicity. The “star” ratings are easy to understand, comparison among plans is all on one page, and ratings are broken down into categories that many people will find important. On the other hand, NCQA’s Health Insurance Plan Rankings include more detail, and that would benefit the people who do choose to spend more time understanding the quality dimensions of the plan. Either approach (the Accreditation Report Card or Plan Rankings) would present valuable information to consumers. The main advantage of either of these approaches over the “Overall Consumer Satisfaction Score” derived from CAHPS is that the memo states there’s little variation in the composite CAHPS score, while the drill-down option there is a bit overwhelming.

Thanks very much for considering our input.
Kurt Hoppe, MD: Mayo Clinic

I had a chance to think about the options presented on Monday.

Is it possible to merge/hybrid two options? In this case, options 3 & 4? I like the features of both options: the ease of the star rating system of option 3 and the “granularity” of option 4.

Thanks for allowing me to participate.
Sue Knudson: HealthPartners

**Background**
Consumers’ Checkbook has proposed five options for reporting health plan quality data in the Exchange for the first open enrollment in October 2013. The options range from reporting no quality results at all to a variety of options that report some degree of quality performance.

**Recommended Approach**
HealthPartners agrees that there is a substantial amount of comparable quality data available from most Qualified Health Plans offered by insurers and the Exchange should make it available to allow consumers to have these data to complement the cost data. We support Option 5 which uses a combination of available data. Specifically, the NCQA Accreditation Report Card and the NCQA Health Insurance Plan Rankings to give consumers access to these widely available measures of Health Plan quality. We support this option since it best meets both the consumers who want the high-level plan rankings and those that want the detailed measure results.

One important consideration for implementing Option 5 is to ensure special attention is paid to not comingle or substitute results by payer type or product. This can be achieved by clearly identifying the payer and product represented and substituting industry nomenclature (i.e. HMO, PPO, etc) with meaningful descriptions for consumers. Also, the exchange should only report data which supports the payer type and products being offered and not inferring performance from other payer types and products offered by a plan. Evidence supports that within a plan, Medicaid and commercial member rates can be markedly different. Therefore, if a plan has results only for Medicaid, it should not be inferred that the rate is representative of commercial performance.

Not all health plans participating in the exchange will have quality performance data representative of all payer and product types they are making available in the Minnesota exchange. However, this should not deter us from presenting the quality performance results that do exist for the vast majority of the market participants. Language should be developed to ensure new market entrants are not disadvantaged by the lack of available data. As discussed at the last meeting, new entrants could come in each year so this situation will be ongoing. Holding back information when it exists is not a recommended strategy.

**Citations**
Potential Methodologies for Constructing a Composite Quality Measure

(March 18, 2013 meeting topic)

Sue Knudson: HealthPartners

Thank you for the opportunity to provide comment as part of the Measurement and Reporting Work Group of the Minnesota Health Insurance Exchange Advisory Committee. Below are my suggestions as they pertain to the Quality Ratings System (QRS) options presented by Consumer’s Checkbook at the meetings on February 25, 2013 and March 18, 2013.

1. Data Availability for Proposed Quality Rating Systems (QRS) Measures and Considerations Related to Sample Sizes and Potential for Small Numbers Issues

The QRS measures should be limited to the HEDIS/CAHPS measures, without modification. These measures are currently part of all health plan operating procedures, are externally audited and highly credible measures used to inform consumers about the quality of a health plan. The other suggested measures are inconsistently used or are not in use in Minnesota and would require significant resource and cost investment to implement for likely little consumer gain.

The eval8 tool is a three hundred plus page request that requires thousands of hours of work to complete for a health plan. This tool is not currently in use by Minnesota health plans and we recommend not requiring use of the full survey across plans. We recommend considering use of a few key measures for the Quality Rating System if it is desired to complement existing measures with plan capability information from this tool.

The existing accreditation levels should be extended to the QRS for the related products (HMO, PPO, etc) and the related payer (Commercial, Medicaid). The option presented by Consumer’s Checkbook to re-sample and reduce sample sizes for the exchange products that would create wider confidence intervals and results in a consumer’s inability to distinguish meaningful differences in plan quality.

Finally, NCQA and URAC accreditation are not inter-changeable or comparable as it relates to quality reporting. Using results from both accrediting organizations will likely cause consumer confusion and frustration when trying to compare the quality of the plans offered in the exchange. While there is some measure overlap, URAC is less intensive and less costly than the NCQA accreditation process.

2. Methods for Creating Composites
As mentioned above, our strong preference is to use NCQA accreditation and Consumer Reports health plan quality ratings. In lieu of this, HealthPartners recommends Method 4 with a performance threshold, rather than the state average for determining achievement. This keeps the measurement relevant to Minnesota plans performance while simultaneously rewarding high performance where it exists. Method 4 addresses some of the key issues in measurement compositing by:

- Scaling individual metrics the same.
- Incorporating significance testing to ensure actual differences in performance are evident.
- Comparing individual measures to meaningful quality thresholds rather than state or national averages. Measures compared to the state average will force “losers”. This is particularly important when performance differences are very small or overall performance is very high.

The “con” articulated for Method 4 states reporting entities with large sample sizes (or in many cases, full populations) are more likely to be significantly different from the average as compared to entities with smaller sample sizes. This should not be seen as a negative attribute of this method. Some entities may report their full population performance rates and thus the confidence intervals will be small, reflecting precise performance.

3. Weighting individual measures and sub-composites
   - HealthPartners prefers an “expert” panel to set measures and sub-composites weights. The panel should ensure weights are balanced so that particular sub-composites do not overwhelm all other sub-composites as part of the overall composite measure. For example, chronic conditions may have significantly more measures than preventive measures. Without smart calibration, the chronic conditions could overwhelm the score. The expert panel should also consider consumers values in determining appropriate weighting for the quality measures.

4. Handling missing data
   - HealthPartners agrees with the approach to require at least half of the sub-composite category measures available for an entity to calculate that entity’s composite quality score. By creating a quality composite with less than half of the available measures, the majority of the score is imputed (or considered average). In addition, this allows for a few measures to overwhelmingly drive the composite score, whereas the score is intended to represent broader quality performance. In the display, health plans that do not have enough measures to create a composite score (at least half of the measures in each sub-composite) would not have an overall composite score and measure results would not be imputed to create one. In these instances, sub-composite and individual measure results should be available for consumer transparency. Simply being transparent and displaying what actual results are available for these health plans.

Last, we would advise the use case testing with consumers focus on information display and informational relevance after appropriate methods are decided.
Karen Schirle: Department of Human Services

- I prefer the average standardized scores on indices. The QRS should differentiate products for consumers, not homogenize results. This approach also allows for significance testing which could be valuable for research and analysis when the HIX and the QRS matures.
- However the QRS processes data to arrive at the composite measures, the results should have some graphical presentation for ease of consumer use.
- In line with my first comment, I think that measures without enough responses should be dropped rather than replaced by an average which will tend to blur distinctions.
- I think there should be some threshold for responses under which composite measures are dropped for an product for measurement integrity. The contractor’s 50% threshold seems as good as any and since they could reference CMS’s use of that threshold for some specific effort, there is an outside benchmark to support that level.
- I like the “subject matter expert” approach to assigning weights to measures and composites.
- The plans’ reference to a threshold for the measures concerns me since consumers need the differentiation to assist in making choices and having a threshold like Lake Woebegon where all the children are above average will not be helpful to consumers unless that threshold has some relationship to objective reality. And the standardized scores will only tell consumers how products’ quality scores relate to each other’s. While I understand the plans’ desire to all appear to be above average or above some quality threshold based on their needs as businesses, a threshold approach will not assist consumers if set at the lowest common denominator.
Allie Coronis: Allina Health

Some additional feedback with regard to the composite and data discussion from our last Measurement and Reporting Workgroup call.

• What method should be used for creating a composite measure?
  
  I would propose to use Method Two. I like that it is a standardized average and that it allows for the most part, statistics on the composites. I feel the significance piece can be addressed by using proper weights for the most part.

  Method Four, with at least 5 scales would be the method that I prefer next best, although it seems that there is a danger that it would mostly benefit and harm the larger providers by volume.

• What approach should the Exchange use in determining weights for individual measures and/or for subcomposites?

  I think this should be multifactorial considering things like Volume for the measure (essentially denominator for those where an N can be determined), Importance (essentially a clinical weight for the measure), and Variation (essentially a determinant representing the type of distribution in a measures). The weights could be determined based on the factors, and I think that it would be helpful if there were a limited number of potential weights. So, if there were 3-5 weights for each category, one could average those weights for an overall weight.

• How should the Quality Rating System address issues associated with missing data?

  I would propose the omission weighting method. If an element of a composite is missing, just consider the metric with those composites available. It would be important to have clear and adequate data explanations. It might also be a good option that if most metrics are missing that the composite is not displayed.

• What method of comparing reporting entities should be used?

  Option #2 (significance testing seems best to me. However, to mitigate the effect of just large hospitals will hit the 95% confidence, I would suggest a number other than this. In fact, I would suggest something like the 5 scale of method #4 of the composite creation. This might be the best way to include some of the lower volume entities.

Thanks for the opportunity to provide additional comment.