CONSUMERS' CHECKBOOK
Center for the Study of Services

CONSIDERATIONS RELATED TO SAMPLE SIZES AND POTENTIAL FOR SMALL NUMBERS ISSUES

Looking forward to the Exchange's future open enrollment periods in years beyond 2013, a set of 74 measures for inclusion in the Quality Rating System (QRS) has been proposed to the Measurement and Reporting Work Group. This memo begins to address the practical issues with implementing the proposed QRS by raising issues associated with timing and sample sizes as well as strategies for addressing sample size issues.

Research conducted for the Minnesota Health Insurance Exchange estimates that the total number of consumers obtaining health coverage through the Exchange may reach 1.2 million consumers by 2016. The Exchange also anticipates a ramp up period during the initial launch and operation of the Exchange as consumers learn about the Exchange and how to use it to obtain health coverage. In addition, in the early years of the Exchange, some Qualified Health Plans (QHPs) may attract a large number of enrollees, while other plans attract few. For both of these reasons, the Exchange must provide for the possibility that smaller enrollment numbers in QHPs will affect the ability of QHP-level information to be publicly reported in the Quality Rating System (QRS). This memo will discuss the timing and sample size requirements for measurement of the predominant proposed measures included in the straw model QRS, the HEDIS/CAHPS measures. Options for addressing small sample issues will also be outlined.

Timing of Measurement

Before measurement can be conducted at the QHP level, time must pass to allow members to have experiences with a specific QHP. Most HEDIS measures, including the HEDIS Consumer Assessment of Healthcare Providers and Systems (CAHPS) enrollee experience survey, require members to be enrolled in a plan for an entire measurement year. HEDIS measurement years correspond with calendar years.

Some measures in the proposed measure set require members to have more than one year of plan enrollment before the member is included in the measure denominator. Examples include:

- *Persistence of Beta-Blocker Treatment After a Heart Attack*
- *Breast Cancer Screening*
- *Cervical Cancer Screening*
- *Prenatal & Postpartum Care*

If Minnesota QHPs start providing service to enrolled members in January 2014, then the earliest that any measures could be collected at the QHP level would be January 2015. Some HEDIS measures could not be collected at the QHP level until January 2016 at the earliest.

Sample Size Required

All HEDIS clinical measures have a minimum sample size of 411. The sample size of 411 specified by NCQA is based on a statistical estimation of providing an 85 percent chance of identifying a five percentage point difference between plans. The calculation of the required sample size is dependent on three factors.

1. The percentage point difference between plans desired to be identifiable
    - NCQA has chosen their sample size to allow the ability to identify a five percentage point difference between plans. If one wishes to have the ability to identify smaller differences between plans, then the sample size must increase. If one is satisfied only identifying differences greater than five percentage points, then the sample size can decrease.
2. The likelihood of identifying the difference
    - Likelihood of identifying the differences between plans is the likelihood that the minimum percentage point difference identifiable will be detected. To increase this likelihood, one must increase sample size.
3. The average score of the measure
    - In setting the sample size at 411, NCQA has made the conservative assumption that scores will average 50 percent. As a score moves closer to 0 percent or 100 percent, a smaller sample size is required to detect the same percentage point difference.
        - The formula used to identify differences between proportions takes into account the standard error of each proportion. [1] As a result, the standard error is at its greatest when a proportion is .5 and it decreases as proportions approach 0 or 1. The implication of this formula is that the confidence interval around the proportion of 50 percent will be greater than the confidence interval around any other proportion given that the sample sizes are the same, thus making differences between proportions that average 50 percent most difficult to identify.

The table on page three illustrates how the required sample size would change as a result of changes in each of the three factors listed above.

[1] The formula for the standard error of a proportion is (p*(1-p)/n)^.5, where p is the proportion and n is the sample size.

	Minimum percentage point difference detectable	Likelihood of identifying difference	Average score	Sample size required of each reporting entity
NCQA standard	5%	85%	50%	411
Alternative standard for detecting difference in score	4%	85%	50%	643
	6%	85%	50%	286
	10%	85%	50%	103
Alternative standard for likelihood of identifying difference	5%	80%	50%	329
	5%	90%	50%	542
	5%	95%	50%	769
Alternative standard for average score	5%	85%	75%	309
	5%	85%	85%	210

The HEDIS CAHPS survey for commercially enrolled plan members has a sample size of 1,100. Commercial carriers average a 35 percent response rate; so carriers collect on average about 385 returned surveys per commercial product line. Respondents are instructed to skip some questions on the survey based on their experiences in the prior twelve months. As a result, some questions are answered by as few as 30 percent of survey respondents. NCQA has a minimum threshold of 100 usable responses for an individual HEDIS CAHPS survey measure to be considered reportable.

Enrollment Required to Reach Sample Size

The ability to reach the desired sample size for a HEDIS measure will depend on the number of plan enrollees that qualify for each specific measure. This varies dramatically by measure based on characteristics of the plan enrollees, as demonstrated by the following examples.

The *Cervical Cancer Screening* measure includes in the denominator all women age 24 to 64 years old (with the option of excluding women who had a hysterectomy with no residual cervix). Looking at commercial plan enrollee distribution, it is reasonable to expect 30 percent of plan enrollees to qualify for this measure. Targeting a sample size of 411 would mean that a QHP with enrollment as low as 1,400 might be able to reach the target.

For measures that require enrollees to be young children, such as *Well-Child Visits in the First 15 Months of Life* where only children that turn 15 months old during the measurement year are included in the denominator, or *Childhood Immunization Status*, where only children that turn two years old during the measurement year are included in the denominator, it is reasonable to expect that as few as one or two percent of QHP enrollees will qualify for the measure. Targeting a sample size of 411 for these measures would mean that a QHP would require total enrollment as large as 41,000 to be able to reach the target.

The denominator for the various diabetes measures is patients 18-75 years of age as of December 31 of the measurement year who had a diagnosis of diabetes (type 1 or type 2). Based on information collected from the CDC website, 6.5 percent of adults in MN were diagnosed with diabetes in 2010. If 70 percent of QHP enrollees are 18-75 years of age, this means that 4.5 percent of total QHP enrollees would qualify for the measure. Targeting a sample size of 411 would mean that a QHP might need to have more than 9,000 persons enrolled to be able to reach the target.

Options for Shortfalls in Sample Size

Based on their enrollment, QHPs may have difficulty reaching the target sample sizes for some of the measures. Options for addressing the shortfall would be:

1. During the initial years of Exchange implementation, require data collection and reporting at the product line level or product line/metal tier level rather than at the QHP level. Assess opportunities for moving to QHP level data collection and reporting over time.
    a. If a carrier is offering multiple QHPs of the same product type within the Exchange, then enrollees across the QHPs could be combined for measurement purposes.
        - This strategy would provide for more consistency in data collection and reporting practices across carriers as compared to other options; however, this would not be a viable option for carriers offering a single QHP within a product type.
2. Require data collection and reporting at the QHP level, using one or more of the strategies below to address small sample size issues.
    a. Establish a lower threshold for reporting for QHPs that are unable to reach the target sample size.
        - Allowing QHPs to report with a smaller sample size would make differences between QHPs less likely to be identified.
    b. Allow QHPs to combine data collected over multiple years.
        - Combining data over multiple years would have the negative effect of dating the results reported to consumers. Results reported for Open Enrollment in October of 2017 would reflect QHP performance from January 2015 through December 2016 partially masking the impact of quality improvement initiatives taken by the plan.
        - Comparability of results between QHPs could also be impacted if some QHPs combine data over multiple years while others do not. General changes in clinical practice or other external factors may raise or lower all QHP scores in a given year. If some QHPs are combining data over multiple years, then the impact of these external factors will be dampened for these QHPs only.

Carriers need to know specific requirements related to quality measurement before a measurement year begins. Measuring at the QHP level requires sampling to occur for each QHP; if data is to be collected at different level of aggregation – such as QHPs within a metal level or QHPs within a product type – carriers need those specifications in advance to measure consistently; therefore, decisions should be made in advance with respect to whether sampling needs to occur at the QHP level or a higher level

of aggregation. It should also be noted that carrier administrative costs increase with the number of samples required.

Discussion Questions:

- How would potential strategies for addressing small numbers issues affect both existing carriers in the individual and small group markets as well as new entrants for purposes of the Exchange?

- What strategy is most appropriate for addressing the potential for small numbers issues during the initial years of Exchange implementation?