NATIONAL QUALITY FORUM

Meeting Summary

MAP Rural Health Technical Expert Panel Web Meeting

The National Quality Forum (NQF) convened a public web meeting for the MAP Rural Health Technical Expert Panel on November 13, 2018.

Welcome, Roll Call, and Review of Meeting Objectives

Ameera Chaudhry, NQF Project Analyst, welcomed participants to the web meeting. Ms. Chaudhry provided opening remarks and conducted a Panel roll call. Karen Johnson, NQF Senior Director, briefly reviewed three meeting objectives: to explore the pros and cons of previously recommended solutions to the low-volume challenge; to discuss additional statistical methods for addressing the low-volume challenge; and to discuss a strategy for drafting the project report.

Discussion of the Pros and Cons of Previously Recommended Solutions for Addressing Low Case-Volume

The TEP discussed the pros and cons of previously recommended solutions for the low case-volume problem.

Selected measures (particularly for P4P programs) that are broadly applicable to large numbers of patients (e.g., screening measures)

Pros	Cons
1. Maximizes statistical payoff and population coverage for limited set of measures and helps to control burden on providers	1. Artificially constrains measures set to common measures (perhaps mostly toward processes rather than outcomes)
2. Selection criterion of "*being broadly applicable to large numbers of patients*" can easily be combined with other criteria used to select measures for particular programs	2. May result in the neglect of other measures that might be important for the rural population
3. Some measures that are broadly applicable to large numbers of patients are particularly relevant to rural populations (e.g., screening, immunizations, diabetes, transitions)	3. May "hold rural areas back" if focus only on things like screening or immunizations
4. Should increase participation of low-volume providers (LVP) in accountability programs	

Pool data across several years (e.g., using three years of data rather than just one year)

Pros	Cons
1. Increases effective sample size and therefore, likely would improve reliability	1. Decreased sensitivity to changes in improvement over time, and results are less timely and therefore less relevant to decision making (by various stakeholders). This seems particularly problematic in the context of programs whose explicit goal is to incentivize such changes. NOTE that empirical investigation would be needed to determine whether there are actually measurable program effects in short time windows, or whether the short-term perspective just encourages misleading over-interpretation of noise.
2. Likely to provide opportunity to assess performance reliably within patient subgroups	2. Different choices of the pooling window can lead to different results [NOTE that this is true of any variation in methods]
3. Should increase participation of low volume providers (LVP) in programs	

Additional Discussion

TEP members agreed that choice of the timeframe would affect measure results. They agreed that a "rolling" timeframe may help address the timeliness of measure results (i.e., calculating measure results more frequently so as to incorporate the most recent information).

Dr. Finucane suggested a variation of this "pooling of years" idea. If one considers use of only one year of data as "no pooling" and use of an arbitrary number of years (e.g., a 3-year timeframe) as "complete" pooling, one could also consider a "partial pooling" approach. Using such an approach, one could increase reliability by combining data across years, but only to the extent that the data suggest changes over time are due to noise rather than signal. In other words, a partial pooling approach would strike a data-driven compromise between no pooling and complete pooling by incorporating provider data from previous years, but down-weighting it relative to data from the current year depending on how much noise there is in the data overall. The algorithm that determines the down-weighting would be consistent across providers, even though the weights across time might vary (e.g., some providers would have zero weights for earlier years but providers with low volume and/or low precision would have higher weights for earlier years).

Mr. Geppert noted that this approach is similar to the concept of "mastery," wherein one could set a reliability threshold and use as much past data as needed to reach that threshold. Dr.

Finucane agreed that the partial pooling approach would be a way to implement the "mastery" approach. She further clarified that the partial pooling approach can be considered, conceptually, as "borrowing strength" for a particular provider from his or her past performance. TEP members agreed that for such an approach, the available "window" of data available would be constant across providers (e.g., three years or five years, etc., whatever would be practical in terms of implementation), although how much of that past data would be used would vary across providers.

Aggregate data from multiple providers (e.g., combining data within regions or networks)

Pros	Cons
1. Increase effective sample size and likely would improve reliability	1. Decreases sensitivity to differences across providers in the same region/network NOTE that empirical analysis would be required to determine whether reliable measurement would be possible at individual provider level.
2. Results in smaller number of provider groups to present (this may be particularly appealing for public reporting)	2. Determining which regions to combine likely would be subjective/arbitrary In addition, the definition and composition of physician groups likely would not always be consistent.
3. Possibility for greater variation in patients for each provider group, which could potentially facilitate subgroup analysis	3. Combining data across providers may still not be enough for small population groups/rare outcomes
	4. Estimates become less useful for guiding quality improvement activities or other types of decision making
	5. Could possibly reduce perceived incentive to providers for QI effort

Additional Discussion

As above, a "partial pooling" approach could be a best-of-both-worlds solution that would increase reliability by combining data across providers, but only to the extent that the data suggest differences across providers are due to noise rather than signal. In this context, "partial pooling" is a synonym for "reliability adjustment," "shrinkage," "borrowing strength," etc.

Combine inpatient and outpatient data for similar measures

Pros	Cons
1. Increased effective sample size and likely would improve reliability and power	1. There is a risk that analysis is dominated by inpatient (or outpatient) treatments, depending on disease types
2. Potentially more representative assessment of performance across the range of patient severity	2. Would eliminate the potential differences between inpatient and outpatient data
	3. For some measures, inpatient and outpatient experience may represent substantively different processes
	4. Performance data would be less relevant to decision making

Additional Discussion

The TEP agreed that this solution would be technically feasible for various topics. They also noted that it might be preferable for certain types of measures or topics, including resource use. For example, combining both emergency department visits and admissions would allow measurement of hospital utilization. Similarly, combining data may be needed to adequately assess certain conditions or treatments (e.g., including both inpatient and outpatient surgeries for breast cancer). While attribution may be an issue, for many measures, the accountable entity likely would be the provider involved in the index encounter. TEP members agreed that the utility of this solution would depend on the performance metrics being considered.

Develop composite measures that expand the number of patients captured by measurement

Pros	Cons
1. Aggregating across multiple measures could reduce sampling variation/increase stability of measure results	1. A small number of "signals" may be masked by a large number of "noises"
2. Could reduce the number of measures being reported (although it likely would not affect the burden of data collection)	2. May mask important differences in key components of care to the extent performance on one individual component is not related to performance on another
3. May better capture care complexity for particular conditions	3. May not be easy to understand how to improve
4. Potential to increase effective sample size	

Additional Discussion

The TEP thought that the type of composite (e.g., reflective vs. clinometric) likely wouldn't matter in terms of this being a viable solution to the low case-volume challenge. Dr. Zaslavsky noted that when combining measures that are relatively uncorrelated at the patient level (e.g., a component measure that assesses care of asthmatic patients, another component that assesses care of diabetes patients, etc.), this not only increases the de facto "denominator" of the measure, but also results in relatively large reductions in sampling variance. One member noted that some composites might also increase the effective size of the numerator, which may also be helpful.

Present confidence intervals, numerator counts, and denominator counts

Pros	Cons
1. Confidence intervals reflect the high uncertainty inherent in the low-volume context	1. Confidence intervals can be inverted to perform null-hypothesis significance tests (NHST) (a misuse of the confidence intervals). This could be a problem if the measure is calculated without using shrinkage. An alternative approach to reflecting uncertainty, without risking the errors inherent in the NHST framework, would be to use the "exceedance probabilities" recommended in the COPSS report.
	2. CI may not really help with decision making.

Additional Discussion

Although TEP members acknowledged potential misuse of confidence interval information, they still agreed that some sort of presentation of uncertainty in measurement would be useful, even though it might be unlikely to change decision making. TEP members also noted that even with reporting of uncertainly, the incentive system itself would have to be well-designed.

Use indicators that do not have a denominator* (e.g., number of infections per month; time since last adverse event)

Pros	Cons
1. There is one less source of uncertainty	1. Not all measures can be accommodated. Proportions (which demand denominators) can be especially informative for some topics
2. Perhaps could be used in accountability applications (e.g., never events)	2. Inability to meaningfully compare across sites. Sensitive to shifting patient populations within a facility over time. Best suited for internal quality improvement purposes
	3. Unclear how to compare performance (or make decisions with respect to performance)

Stratify providers so that performance results are compared only among similar groups (i.e., comparing "like to like")

Pros	Cons
1. Stratification is a common (and highly useful) tool for eliminating variation	1. If rural providers are only compared to other rural providers, then we would lose the ability to determine whether rural providers are systematically higher or lower performing than their nonrural counterparts.
2. Statistical tools are well developed for stratified analysis	2. When multiple possibilities are present, choosing the proper stratification variable(s) can be challenging. There is a risk of "over stratification." (There are some principles that can be applied to assess the appropriateness of various stratifications. However, this will depend on the objectives that are supposed to be served by the presentation of the statistics; can't really get far with stratification objectives without those.)
3. Potentially accounts for unobserved patient heterogeneity	3. Does not increase effective sample size and thus doesn't really address the low case-volume problem (instead, it addresses fairness in comparisons)

Consider measures that reflect the wellness of the community (i.e., population-based measures)

Pros	Cons
1. Increases sample sizes	1. Determination of how to define community is subjective.
	2. Healthcare determines a [small] part of community health.
	3. Estimates are less useful for guiding quality improvement activities or other types of decision making.
	4. Does not allow for accountability at an individual provider level

Reconsider exclusions for existing measures

Pros	Cons
1. Can increase the effective sample size	1. Would introduce complexity in specifications if exclusions had been used as a way to ensure patient or treatment homogeneity
	2. Care would have to be taken not to disproportionately affect/penalize rural providers (e.g., excluding transfer patients in readmission measures actually "rewards" rural providers who do the right thing when transferring patients).

Consider measures constructed using continuous variables

Pros	Cons
1. More informative; statistically more powerful	1. Not as easily interpretable as categorical variables

Consider ratio measures

Pros	Cons
1. May be more informative than dichotomous "any/none" measure.	1. The denominator may be related to quality (e.g., number of days on a central line)
	2. Likely would not actually be adding a lot of information and therefore not really increasing the reliability of the measure even though increasing the denominator

Discussion of Additional Statistical Methods for Addressing Low Case-Volume

The TEP discussed three additional statistical methods, as follows.

Pool data across providers and across time

To make the best use of all available information, consider combining the ideas of pooling data across years and across providers. Under such an approach, a provider's score would be informed both by data from that provider in previous years and also by data from other providers in the same region/network. The weights placed on these two sources of outside information would be determined entirely by the data.

The cons of this approach include lack of transparency (i.e., the approach is very complicated and thus not easily understood and likely difficult to implement; the likelihood of high computational burden; and the need for a professional statistician to develop).

NOTE that this idea could be expanded further by "borrowing strength" from measures that are related to the measure or from other types of information. One difficulty is that other information may not be freely known (some may be "private" for providers while other information may be private for implementers or sponsors of the measure). However, this isn't insurmountable (e.g., a tool could be developed that provides some data and invites input of other data).

TEP members noted that this approach is known in terms of methods, but the challenge lies in implementing the approach. Contractors (for example, to CMS) would have to have the technical knowledge and the computational power to implement the approach. They would also need a PhD statistician who could develop the models.

Non-parametric alternative: BART (Bayesian Nonparametric Modeling)

This approach has recently been recommended for teacher performance measurement in the education setting, which is statistically a very similar problem to hospital performance measurement. This approach would allow developers to relax assumptions used in more typical (parametric) approaches (e.g., additivity and linearity) and therefore obtain better risk adjustment. Dr. Zaslavsky, however, suggested that use of assumptions is needed when case-volume is low.

Hierarchical modeling

The hierarchical modeling approach accounts for the inherent nested structure of the data and supports both risk adjustment and stabilizing hospital-specific performance estimates by shrinking estimates toward an appropriate target. It is critical, however, that the shrinkage target be meaningful. The "usual" shrinkage target is the national mean, but other targets may be appropriate (e.g., something that is related to procedure volume). Alternatively, measure developers could carefully consider use of hospital attributes in producing shrinkage targets as opposed to shrinking to an overall mean. Use of this type of approach should depend on the intended use of the measure (e.g., quality improvement versus public reporting).

TEP members also briefly discussed use of data known at the start of the performance period versus data during the performance period; however, they did not come to complete consensus on this issue.

Strategies for Drafting the Project Report

NQF staff has begun to outline the report and has made it available as a Google document on the TEP SharePoint page. The TEP agreed that NQF should set up the framework but also provide assignments and deadlines for TEP input. TEP members will use track changes when adding to the document.

NQF Member and Public Comment

NQF staff opened the call to allow for public comment. No public comments were offered.

Next Steps

The next TEP call will be held on November 30, 2018.