Physician Compare Benchmark Technical Expert Panel Summary

Benchmarks can help to ensure that the quality data published on Physician Compare (PC) is accurately interpreted and can allow consumers to more easily evaluate the information by providing a point of comparison. The 2015 Physician Fee Schedule (PFS) proposed rule proposed a benchmarking methodology for public reporting on PC. The proposed methodology aligned with the methodology currently used under the Medicare Shared Savings Program (Share Savings Program). Details on this methodology can be found on CMS.gov at http://cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/MSSP-QM-Benchmarks.pdf. Based on concerns raised during the public comment period, the proposed benchmarking methodology was not finalized. Instead, CMS noted additional work with stakeholders would be done and other programs' methodologies would be closely evaluated in order to work toward developing a more appropriate benchmark methodology for PC.

The PC support team employed a staged approach to engage stakeholders in the development of a publicly reported benchmark. This process includes the following five phases:

- Phase One: Exploratory conversations with 2015 PFS proposed rule commenters
- Phase Two: Internal CMS stakeholder discussion at the Physician Compare January JAD
- Phase Three: Informal call for input via e-mail
- Phase Four: Public webinars
- Phase Five: Technical Expert Panel (TEP)

Considering all feedback collected during four previous stages of the stakeholder outreach process, the PC support team presented salient issues for discussion with interested members of the Physician Compare TEP during a 2-hour focused TEP meeting held in early March 2015. The PC support team's presentation was supplemented with a summary document of issues and considerations provided for review in advance of the TEP.

Following a short presentation by the PC support team, the considerations for a publicly reported benchmark were discussed with the TEP.

Item-Level Benchmark

Understanding stakeholder concerns with a composite-level benchmark, TEP members agreed that an item-level benchmark is the best approach for a publicly reported benchmark at this point. One TEP member noted that creating a composite with individual items that were not developed as composites is like "mixing apple and oranges." This echoed previous feedback explaining that a composite must have internal consistency and reliability, and therefore all composite items should be developed specifically to relate to one another. TEP members also agreed with stakeholders and CMS that a composite-level benchmark should be the ultimate goal. The TEP emphasized that CMS should work with measure developers to ensure composite-level measures relevant for public reporting are a high priority for development.

Benchmark Stratification

The benchmark stratification discussion primarily focused on issues related to patient population and measure availability. TEP members expressed concern that health care professionals (HCPs) may be penalized for treating high risk or complex populations and there was some discussion of integrating the complexity of the patient population into a benchmark methodology. TEP members expressed that incorporating patient population into the benchmark may create a more granular understanding of the quality of care based on both the HCP's specialty and population served. Another TEP member suggested using a methodology based on improvement goals to address complex patient populations. Improvement goals, as compared to performance goals, could encourage improvement from HCPs with lower performances and ensure those HCPs treating high risk or complex populations are not compared to those treating generally healthy populations. Some TEP members supported the usefulness of improvement goals, but others expressed that consumers may misinterpret improvement over time as a performance goal. While a number of TEP members agreed with importance of improvement goals, particularly in the context of payment or incentive, they did not consider improvement goals suitable for public reporting. Ultimately, the TEP agreed that issues around patient population were best addressed through risk adjustment. The TEP strongly recommended that CMS encourage measure developers to build in risk adjustment into the future development of composites for public reporting.

In addition, one TEP member noted that certain specialties do not have relevant measures to report under PQRS, and thus may not have a benchmark to publicly report on Physician Compare. The TEP member was concerned that comparisons between HCPs may not draw accurate interpretations by consumers and that consumers perceive missing measures as negative. The idea of stratifying by specialty was therefore discussed. In addition, to considering stratification by specialty, the TEP member recommended that Physician Compare include language stating particular specialties do not have adequate measures to report in the early phases of public reporting, in particular. Ultimately, the TEP appreciated that stratifying by specialty was not feasible in the early stages of public reporting as the population of reports is not large enough to produce meaningful comparisons. But, the TEP did encourage CMS to include language about available measures. CMS is now considering the feasibility of adding such language for the next measure release.

Benchmark Comparison

The majority of the TEP focused on the benchmark comparison. TEP members expressed concern that both comparative averages and performance goal approaches for a publicly reported benchmark would face issues of tight ceiling effects. One TEP member explained that a comparative average approach would lead to high averages, where CMS would have to decide to either give everyone 5 stars or attempt to discriminate star ratings off of small and potentially insignificant differences.

In addition, TEP members noted that the majority of HCPs would eventually meet the set goal when using a fixed performance goal approach. In this example, CMS would have to decide if all reporters who met the goal would receive 5 stars, or if reporters would receive 3 stars for what was labeled as a goal threshold, for instance. TEP members explained that this is likely with process measures, as these are easier to improve on. Outcome measures are inherently more difficult to achieve high performance rates on, so ceiling effects would not be as much of an issue. In the early stages of public reporting, however, there was concern around fixed performance goals for this reason, and because there was no agreed upon goal for the available PQRS measures.

The TEP evaluated the limitations of the measures currently available, discussed the reality that each HCP and group practice can report very different measures based on their scope of care and approach to PQRS reporting requirements, and noted that an ideal methodology could work at the item-level in the early stages of reporting as well as at the composite-level as those measures become available. Based on these discussions,

several TEP members ultimately recommended looking at the Achievable Benchmark of Care (ABC™). This methodology was also previously recommended during the open call for input, a previous phase of the Benchmark Outreach process. TEP members explained the ABC™ as a "best in class approach" that is ultimately a hybrid of the comparative average and performance goal approaches, where current performance can be captured at the high end to create a "goal." The approach uses an algorithm that focuses on high performance and adjusts for small numbers of cases so that rare events will be considered without overly influencing the benchmark. TEP members sent several peer reviewed articles to the PC support team demonstrating that the methodology is research-based, data driven, and well-defined. Given the evidence base supporting the methodology and the fact the methodology could accommodate the many factors influencing the measures available for public reporting on Physician Compare, the TEP agreed this was a beneficial approach.

Star Rating Display

During the TEP, more attention was given to how to display the stars than during previous rounds of stakeholder outreach. TEP members believe it is important to highlight which measures are more difficult to achieve high scores and the TEP recommended several solutions for consideration. One TEP member recommended developing a ranking list of measures, where measures are grouped by difficulty. The first grouping could be the easiest to achieve and would be labeled as routine process measures that are readily achieved by most reporters. Each subsequent measure grouping would be more and more difficult to achieve. As progress is made over time, the "difficult measures" may move to easier levels of achievement, demonstrating overall improvement to consumers. This could also allow "reach" measures to be posted and understood by consumers as more difficult to achieve. Similarly, one TEP member recommended using different colors to signify measure difficulty and to communicate to consumers when reporters were achieving well on the most difficult measures. The most popular recommended color scheme was gold, silver, and bronze, with gold being the most difficult to achieve and bronze being the least difficult. Ultimately, this was considered an idea worthy of further discussion.

Summary of Findings

The TEP both echoed findings found during previous phases of outreach and brought in new ideas for moving forward. The findings below summarize the TEP discussion.

- An item-level benchmark is more desirable than a composite-level benchmark given the available PQRS measures, but composite-level benchmarks should be the ultimate goal in the long term.
- TEP members expressed concerns about complex patient populations and measure availability.
- The ABC™ methodology was strongly recommended as it would offer a hybrid approach of the performance goal and comparative average methodologies that is evidence based, data driven, and sets a high but achievable bar for quality.
    - The methodology would ensure more stars are higher quality.
    - The methodology could help ensure meaningful differences between star ratings.
- The benchmark display could differentiate by measure difficulty. TEP members recommended either ranking measures in list form or color-coding stars by measure difficulty level.