2018 Quality Measure Benchmarks Overview

What Are Quality Measure Benchmarks?

When a clinician or group submits measures for the Merit-based Incentive Payment System (MIPS) quality performance category, each measure is assessed against its benchmark to determine how many points the measure earns. In program year 2018, a clinician or group can receive anywhere from 3 to 10 points for each MIPS measure (not including any bonus points) that meets the data completeness standards and case minimum requirements through Benchmarks. Benchmarks are specific to the type of submission mechanism: Qualified Clinical Data Registries (QCDRs), Qualified Registries, Electronic Health Records (EHRs), CMS Web Interface, CMS-approved survey vendor (for the Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys), and Claims. In order to measure performance that is comparable across the spectrum of performance, benchmarks are established using historical data. Benchmarks (either historical or performance period benchmarks) are based on actual performance data from 2016 that was submitted to the Physician Quality Reporting System (PQRS) in 2017, except for the CAHPS surveys. For 2017 CAHPS for MIPS, the benchmarks were based on two sets of surveys: 2015 CAHPS for PQRS and CAHPS for Accountable Care Organizations (ACOs). However, for the 2018 CAHPS for MIPS, we have not yet established benchmarks. The CMS Web Interface uses benchmarks from the Medicare Shared Savings Program.

How Are Benchmarks Displayed?

Each benchmark is presented in terms of deciles. Points will be awarded within each decile (see Table 1). Clinicians who receive a score in the first or second decile will receive 3 points. Clinicians who are in the 3rd decile will receive somewhere between 3 and 3.9 points depending on their exact position in the decile, and clinicians in higher deciles will receive a corresponding number of points. For example, if a clinician submits performance data of 83% on a non-inverse measure, and the 5th decile begins at 72% and the 6th decile begins at 85%, then the clinician will receive between 5 and 5.9 points because 83% is in the 5th decile. For inverse measures where a higher performance is seen by a lower number on the performance score, the scores are reversed in the benchmark deciles.

Historical Benchmark Inclusion Criteria

Benchmarks are established using historical data, if available, which is the performance data submitted (during applicable submission period) by individuals and groups two years prior the start of the applicable performance year for which historical benchmarks are established. With the implementation of MIPS starting in 2017, the availability of historical data pertaining to MIPS for purposes of establishing historical benchmarks will not be accessible until 2019. To address this issue, the utilization of historical data submitted by individuals and groups under PQRS accommodates the time lapse for MIPS to generate historical data while establishing historical benchmarks for quality measures under MIPS. For the 2018 performance year, the historical

data assessed to establish historical benchmarks for MIPS quality measures is based on 2016 PQRS performance data reported by individuals and groups during the applicable submission period in 2017.

In order for PQRS performance data to be comparable and applicable to MIPS for purposes of establishing historical benchmarks, the eligibility criteria pertaining to MIPS participation was applied to the PQRS data. Once the historical PQRS performance data submitted by individuals and groups in 2016 was captured, the MIPS eligibility criteria was applied to such data in order to calculate historical benchmarks for MIPS quality measures based on the performance of individuals and groups that meet MIPS eligibility. For historical PQRS performance data to be used in calculating historical benchmarks, the data reported reflects submissions by the following:

- Individual clinicians who:
  - Are identified as a clinician type under the definition of a MIPS eligible clinician;
  - Are not newly Medicare-enrolled; and
  - Exceed the low-volume threshold established for the 2018 performance year (have more than $90,000 in Medicare Part B allowed charges and provides care to more than 200 Part B-enrolled Medicare beneficiaries) as an individual.
- Groups that:
  - Have at least one clinician who is identified as a clinician type under the definition of a MIPS eligible clinician; and
  - Exceed the low-volume threshold established for the 2018 performance period (have more than $90,000 in Medicare Part B allowed charges and provides care to more than 200 Part B-enrolled Medicare beneficiaries) as a collective group.

Comparable Alternative Payment Model (APM) data was included when possible. Benchmarks were created if there were at least 20 reporting clinicians or groups that met the criteria for contributing to the benchmark, including meeting the minimum case size (which is generally 20 patients), meeting the data completeness criteria (60% reporting rate), and having performance greater than 0 percent (less than 100 percent for inverse measures).

What If A Quality Measure Does Not Have A Historical Benchmark?

For measures with no historical benchmark, MIPS will attempt to calculate benchmarks based on 2018 performance data. The same MIPS eligibility criteria listed above will be applied prior to establishing performance period benchmarks for measures. If no historical benchmark exists and no benchmark can be calculated, then the measure will receive 3 points as long as data completeness has been met. In the list of measure benchmarks, measures without historical benchmarks have no data provide in the decile fields.

Benchmark Descriptions

Each benchmark has the following information:

- Measure name and ID

- Submission type (EHR, QCDR/Registry, Claims)
- Measure type (e.g., outcome, process,)
- Whether or not a benchmark could be calculated for that measure/submission mechanism
- Range of performance rates for each decile to help identify how many points the clinician earns for that measure
- Whether the benchmark is topped out (topped out means the measure is not showing much variability and may have different scoring in future years)
- Whether the measure will receive special scoring in a topped out status (specially scored measures will receive 7 points)
- Whether the measure was topped out in program Year 2017
- Whether the measure was topped out in program Year 2018

Table 1: Using Data Benchmarks to Determine Points (Non-Inverse Measures)*

Decile	Number of Points Assigned for the 2018 MIPS Performance Period
Below Decile 3	3 points
Decile 3	3-3.9 points
Decile 4	4-4.9 points
Decile 5	5-5.9 points
Decile 6	6-6.9 points
Decile 7	7-7.9 points
Decile 8	8-8.9 points
Decile 9	9-9.9 points
Decile 10	10 points

**For inverse measures, the order would be reversed. Where Decile 1 starts with the highest value and Decile 10 has the lowest value*.

Historical Benchmarks with Less Than Ten Deciles

By using historical measure performance data, some benchmarks across one or more submission mechanisms were identified with maximum rates (i.e. 100%) without utilizing all ten deciles. These benchmarks are identifiable when the deciles from three to nine are not populated while the tenth decile is identified at 100%. This is evident in inverse measures as well. Deciles that are not populated indicate that the historical benchmark analysis identified that between 10% to 60% or more of the clinicians performed at the maximum achievable performance rate. For example, in the benchmark for Measure #117 presented below, historical benchmarking identified that the top 40% of clinicians performed at the maximum rate. Therefore, clinicians using this submission mechanism that performed above the 6th decile would receive a maximum performance score of 10 points.

Table 2: Example of a Measure Benchmark with Less than Ten Deciles

Measure Name	Measure ID	Submission Method	Measure Type	Benchmark								
					Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Diabetes: Eye Exam	117	Registry/ QCDR	Process	Y	80.81-92.85	92.86 - 96.53	96.54 - 98.70	98.71 - 99.99	- -	- -	- -	100

Special Considerations

Historical Benchmarks for CMS Web Interface Reporters

For the CMS Web Interface, the benchmarks are the same as the 2018 Medicare Shared Savings Program performance benchmarks. While the benchmarks are the same, the scoring will be adjusted to be consistent with other MIPS measures. In order to align with the Medicare Shared Savings Program not including benchmarks below the 30th percentile (which is the start of the 4th decile), any value below the 30th percentile will receive 3 points. However, if performance is above the 30th percentile, then scoring will be the same as other measures. A link to the MSSP Benchmarks will be provided in early 2018 when they are publicly available.

Benchmarks for Consumer Assessment of Healthcare Providers & Systems (CAHPS) Reporters

For CAHPS for MIPS, 2018 benchmarks will be available for each summary survey measure (SSM). However, the CAHPS for MIPS benchmarks have not yet been established because a revised survey will be implemented in 2018. Therefore, we will calculate benchmarks for 2018 based on 2018 performance data. A range of 3 to 10 points are assigned to each SSM by comparing performance to the benchmark (similar to other measures). The final CAHPS for MIPS score will be the average number of points across all scored SSMs. The CAHPS for MIPS benchmark file will be provided when it is available at the end of the performance period.

Historical Benchmarks for the All-Cause Hospital Readmission Measure

The percentile- and decile- level benchmarks for the all-cause hospital readmission (ACR) measure will be created using the 2016 PQRS/Value Modifier Program Year data. The ACR measure result for a group will be included in creating the benchmark if the group was a TIN with at least 16 clinicians or was a Medicare Shared Savings Program ACO participant TIN, the group had Part B charges greater than $90,000 and billed those services to more than 200 Medicare Part B beneficiaries and was not otherwise excluded and the group met the case minimum of at least 200 cases for the measure. The 2018 ACR MIPS benchmarks will be provided in early 2018.

Historical Benchmarks for Topped-Out Measures

For each process measure, a measure is topped out if the median performance rate is 95% or higher (non-inverse measure) or is 5% or lower (inverse measures). For each non-process measure, a measure is topped out if the truncated coefficient of variation (TCV) is less than 0.10 and the 75th and 95th percentiles are within 2 standard errors. The status of topped-out is identified in the benchmark file that accompanies this fact sheet. For further understanding on how topped out measures are scored, please review the MIPS Scoring Guide.

Benchmarks for Multi-Strata Measures

Some measures have more than one numerator and denominator, or stratum, used to calculate overall performance. These multi-strata measures usually employ the average or weighted average of each numerator and denominator combination (i.e. the stratums are combined). However, in some measures there are specified stratum identified as the primary stratum for a performance rate to use in calculating benchmarks. The list of multi-strata measures and performance processes are provided as a tab in the measure benchmark file package and will be updated and posted in 2018.