

Quality Measure Benchmarks Overview

What Are Quality Measure Benchmarks?

When a clinician submits measures for the MIPS Quality Performance Category, each measure is assessed against its benchmarks to determine how many points the measure earns. In 2017, A clinician can receive anywhere from 3 to 10 points for each measure (not including any bonus points). Benchmarks are specific to the type of submission mechanism: EHRs, QCDRs/Registries, CAHPS and claims. These historic benchmarks are based on actual performance data submitted to PQRS in 2015, except for CAHPS. For CAHPS, the benchmarks are based on two sets of surveys: 2015 CAHPS for PQRS and CAHPS for ACOs. Submissions via CMS Web Interface will use benchmarks from the 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 positive performance is seen by a lower number on the performance score, the scores are reversed in the benchmark deciles.

Historic Benchmark Inclusion Criteria

Eligible performance inclusion criteria were applied prior to establishing benchmarks for measures. The inclusion criteria required performance data to be derived from a type of clinician or group in the list of MIPS eligible clinicians that is not otherwise excluded and to have reported on at least 20 eligible instances and meeting data completeness.

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

For measures with no historic benchmark, MIPS will attempt to calculate benchmarks based on 2017 performance data. Benchmarks are created if there are at least 20 reporting clinicians or groups that meet the criteria for contributing to the benchmark, including meeting the minimum case size (which is generally 20 patients), meeting the data completeness criteria (50% reporting rate), and having performance greater than 0 percent (less than 100 percent for inverse measures). If no historic benchmark exists and no benchmark can be calculated, then the measure will receive 3 points. In the list of measure benchmarks, measures without historic

benchmarks are listed at the bottom of the table. The benchmark calculations for the 2017 performance year used data that was submitted for PQRS in 2015 by clinicians that were a type eligible for MIPS and were not newly enrolled in 2015, or groups with at least 1 such clinician. Comparable APM data is included when possible.

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)

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

Decile	Number of Points Assigned for the 2017 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.*

Historic 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 eligible clinicians performed at the maximum achievable performance rate. For example, in the benchmark for Measure #117 presented below, historic benchmarking identified that the top 20% of clinicians performed at the maximum rate. Therefore, clinicians using this submission mechanism that performed above the 8th 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	EHR	Process	Y	50.57-80.68	80.69-90.05	90.06 - 94.11	94.12 - 96.66	96.67-98.57	98.58 - 99.99	--	100

Special Considerations

Historical Benchmark for Web Interface Reporters

For the CMS Web Interface, the benchmarks are the same as the 2017 Shared Savings Program performance benchmarks. Click [here](#) for details on the Shared Savings Program benchmarks, which are listed in appendix A. While the benchmarks are the same, the scoring will be adjusted to be consistent with other MIPS measures. Because the Shared Savings Program does not post 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. If the 5th decile (p40) begins at 72% and the 6th decile (p50) begins at 85%, then the clinician with a score of 83% would receive between 5 and 5.9 points.

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

For CAHPS, benchmarks are available for each summary survey measure (SSM). From 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.

Historical Benchmarks for the All-Cause Hospital Readmission Measure

The percentile- and decile- level benchmarks for the all-cause hospital readmission (ACR) measure were created using the 2015 PQRS/Value Modifier Program Year data. The ACR measure result for a group was included in creating the benchmark if the group was a TIN with at least 16 clinicians or was an ACO participating in the Shared Savings Program, the group had Part B charges greater than \$30,000 and billed those services to more than 100 Medicare Part B beneficiaries and was not otherwise excluded and the group met the case minimum of at least 200 cases for the measure.

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.

Benchmarks for Multi-Strata Measures

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

Table 3: Example of a Measure with Multiple Strata

Measure Title	Measure Number			Overall Performance Rate 2017	Number of Performance Rates 2017
	CMS	NQF	Quality Number		
Coronary Artery Disease (CAD): Beta-Blocker Therapy – Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF < 40%)	145v5	0070	007	Weighted Average	2