Quality Payment PROGRAM

2023 Merit-based Incentive Payment System (MIPS) Quality Measure Benchmarks Overview

Purpose: This resource provides an overview of how we establish MIPS quality measure benchmarks, how benchmarks are used for scoring, and the information in the 2023 Quality Benchmarks and 2023 Multi-Performance Rate Measures files.

Table of Contents

What Are Quality Measure Benchmarks?

Quality measure benchmarks are the point of comparison we use to score the measures you submit. When you submit measures for the MIPS quality performance category, your performance on each measure is assessed against its benchmark to determine how many points the measure earns.

We compare your performance on the measure against its benchmark.

Beginning with the 2023 performance period,

- We assign anywhere from **1 to 10 achievement points** for each MIPS measure based on this comparison, provided that data completeness and case minimum requirements are met.
- Measures without an available historical or performance period benchmark will receive **0 points**, even when data completeness and case minimum requirements are met.
    - Small practices will continue to receive 3 points.

Original Posting: 1/25/2023

How Are Benchmarks Established?

We establish benchmarks specific to each collection type: Qualified Clinical Data Registry (QCDR) measures, MIPS clinical quality measures (MIPS CQMs), electronic clinical quality measures (eCQMs), CMS Web Interface measures, the Consumer Assessment of Healthcare Providers and Systems (CAHPS) for MIPS Survey measures, and Part B claims measures.

> **Did you know?**
>
> Because benchmarks are specific to collection type, a measure reported as an eCQM will be compared to a different benchmark than the same measure reported as a MIPS CQM.

eCQMs, MIPS CQMs, QCDR Measures, and Medicare Part B Claims Measures

Whenever possible, we use historical data to establish benchmarks. Historical benchmarks for the 2023 performance period for eCQMs, MIPS CQMs, QCDR measures, and Medicare Part B claims measures are based on actual performance data that were submitted to the Quality Payment Program (QPP) for the 2021 performance period. We won't use data submitted for measures that were suppressed in the 2021 performance period to create historical benchmarks for those measures in the 2023 performance period.

To establish a historical benchmark:

- The 2021 and 2023 measure specifications must be comparable (no significant changes to the measure between 2021 and 2023).
- 20 instances of the measure must be reported through the same collection type by individual clinicians, groups, virtual groups, and/or Alternative Payment Model (APM) Entities AND
    - The clinician, group, or virtual group was eligible for MIPS in 2021 (no changes to low-volume threshold for the 2023 performance year), AND
    - The measure met performance year 2023 data completeness (70%) and case minimum requirements (20 cases), AND
    - The measure had a performance rate greater than 0% (or less than 100% for inverse measures).

CMS Web Interface Measures (Medicare Shared Savings Program Only)

We use benchmarks from the Medicare Shared Savings Program (Shared Savings Program) to assess and score CMS Web Interface measures. These benchmarks will be available on the QPP Resource Library. **Reminder:** Beginning with the 2023 performance period, CMS Web Interface measures are available only to Shared Savings Program Accountable Care Organizations (ACOs) reporting the APM Performance Pathway (APP).

Original Posting: 1/25/2023

CAHPS for MIPS Survey Measure

We established a benchmark for each scored summary survey measure (SSM) in the CAHPS for MIPS Survey measure. (Refer to the 2023 CAHPS for MIPS Benchmarks file in the 2023 Quality Benchmarks [ZIP]). These benchmarks were calculated using historical data from the 2021 performance period. A range of 1 to 10 points will be assigned to each SSM by comparing performance against the benchmark (similar to other measures). The final CAHPS for MIPS Survey score will be the average number of points across all scored SSMs.

Administrative Claims Measures

There are 4 administrative claims measures available in the 2023 performance period:

- **Clinician and Clinician Group Risk-standardized Hospital Admission Rates for Patients with Multiple Chronic Conditions**
- **Hospital-Wide, 30-Day, All-Cause Unplanned Readmission Rate for MIPS Groups**
- **Risk-Standardized Complication Rate Following Elective Primary Total Hip Arthroplasty and/or Total Knee Arthroplasty for MIPS**
- **NEW in 2023: Risk-Standardized Acute Cardiovascular-Related Hospital Admission Rates for Patients with Heart Failure under the Merit-based Incentive Payment System**

Beginning with the 2023 performance period, we'll score administrative claims measures exclusively against performance period benchmarks.

How Are Results Displayed in the Benchmark File?

Each benchmark is presented in terms of deciles; the benchmark file displays Deciles 1 – 10. Table 1 identifies the range of points generally available for the measure, based on which decile your performance rate falls into.

Exception: Measures that are topped out for 2 consecutive years are capped at 7 achievement points, even if your performance rate falls in Deciles 7 – 10. The benchmark file still displays values for Deciles 7 – 10 even though the measure can't earn more than 7 achievement points.

Original Posting: 1/25/2023

Did you know?

The 2023 benchmark file also reflects the **flat benchmarks*** finalized through previous rulemaking for **Measures 001 and 236**.

- **Measure 001:** We established **flat benchmarks** for the **MIPS CQM collection type**.
    - We established a performance-based benchmark for the eCQM collection type.
    - There's no historical benchmark for the Medicare Part B collection type, which was suppressed in the 2021 performance year.
- **Measure 236:** We established flat benchmarks for the **MIPS CQM and Medicare Part B claims measure collection types.**
    - We established a performance-based benchmark for the eCQM collection type.

*Flat benchmarks are applied to collection types where the top decile for a historical benchmark is greater than 90% (or less than 10% for inverse measures).

Note: For **inverse measures,** better performance is indicated by a lower performance rate. This is reflected in the benchmark file, where lower performance rates are found in higher deciles.

Original Posting: 1/25/2023

Table 1: Using Benchmarks to Determine Achievement Points for Measures that Meet Data Completeness and Case Minimum Requirements

Decile	Number of Points Assigned for the 2023 Performance Period
No benchmark (historical or performance period)	0 points (small practices will continue to receive 3 points)
Decile 1	1 – 1.9points
Decile 2	2 – 2.9 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

Historical Benchmarks with Less than 10 Deciles

Some benchmarks don't include a range of performance rates for every decile. This occurs when a large percentage of clinicians in the historical benchmark data set have the maximum achievable performance rate. These benchmarks are identifiable when one or more of the deciles between Decile 1 and Decile 9 display "--" while Decile 10 is identified at 100% (or 0% for inverse measures). The higher the percentage of individual c1linicians, groups, and virtual groups that reach the maximum achievable performance rate, the more deciles that will show a value of "--".

For example, in the benchmark results for the Diabetes: Eye Exam measure (Measure ID 117, MIPS CQM) presented in Table 2, historical benchmarking identified that the top 50% of clinicians performed at the maximum rate. Therefore, clinicians submitting through this collection type who performed above the 5th decile would receive the maximum performance score of 10 points.

Table 2: Example of a Measure Benchmark with Less than 10 Deciles

Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
1.30 - 48.28	48.29 - 95.67	95.68 - 99.03	99.04 - 99.73	99.74 - 99.99	--	--	--	--	100.00

Original Posting: 1/25/2023

Did you know?

The **Scoring Examples** tab of the 2023 Benchmark file provides examples for various scoring scenarios.

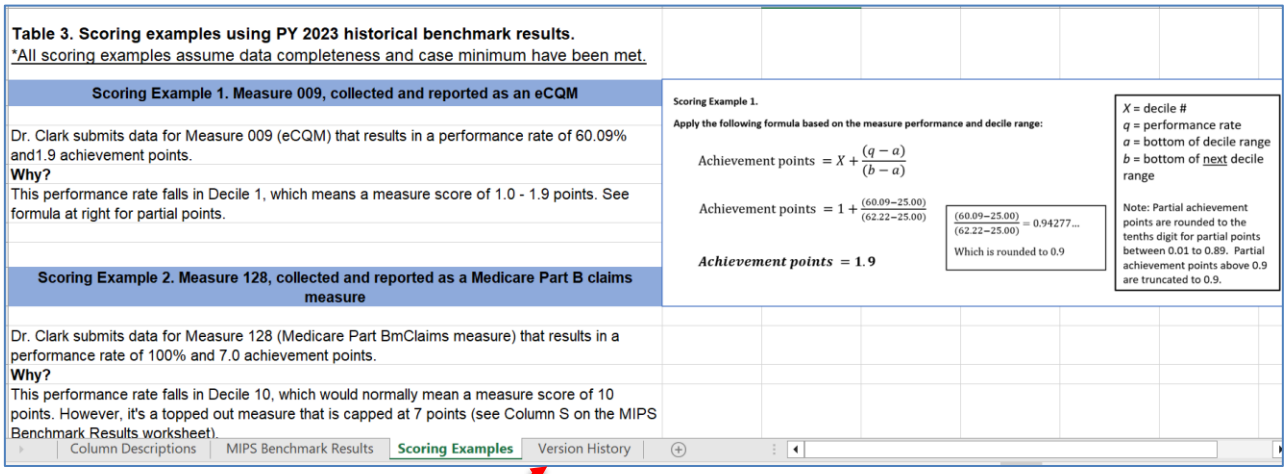


Table 3. Scoring examples using PY 2023 historical benchmark results.
*All scoring examples assume data completeness and case minimum have been met.

Scoring Example 1. Measure 009, collected and reported as an eCQM
Dr. Clark submits data for Measure 009 (eCQM) that results in a performance rate of 60.09% and1.9 achievement points. **Why?** This performance rate falls in Decile 1, which means a measure score of 1.0 - 1.9 points. See formula at right for partial points.
Scoring Example 2. Measure 128, collected and reported as a Medicare Part B claims measure
Dr. Clark submits data for Measure 128 (Medicare Part BmClaims measure) that results in a performance rate of 100% and 7.0 achievement points. **Why?** This performance rate falls in Decile 10, which would normally mean a measure score of 10 points. However, it's a topped out measure that is capped at 7 points (see Column S on the MIPS Benchmark Results worksheet).

Scoring Example 1.

Apply the following formula based on the measure performance and decile range:

$$\text{Achievement points} = X + \frac{(q-a)}{(b-a)}$$

$$\text{Achievement points} = 1 + \frac{(60.09-25.00)}{(62.22-25.00)}$$

$$\frac{(60.09-25.00)}{(62.22-25.00)} = 0.94277\ldots$$

Which is rounded to 0.9

$$\textit{Achievement points} = 1.9$$

X = decile #
q = performance rate
a = bottom of decile range
b = bottom of next decile range

Note: Partial achievement points are rounded to the tenths digit for partial points between 0.01 to 0.89. Partial achievement points above 0.9 are truncated to 0.9.

Column Descriptions | MIPS Benchmark Results | Scoring Examples | Version History

You can also check the **Version History** tab for information about changes made to the benchmark file during the performance period.

What If a Quality Measure Doesn't Have a Historical Benchmark?

If a quality measure or collection type doesn't have a historical benchmark, we'll attempt to calculate benchmarks based on data submitted for the 2023 performance period. We can establish performance period benchmarks when at least 20 instances of the measure are reported through the same collection type and meet data completeness and case minimum requirements and have a performance rate greater than 0% (or less than 100% for inverse measures).

Performance period benchmarks will be established using data submitted by individual clinicians, groups, and virtual groups that are eligible for MIPS in the 2023 performance period.

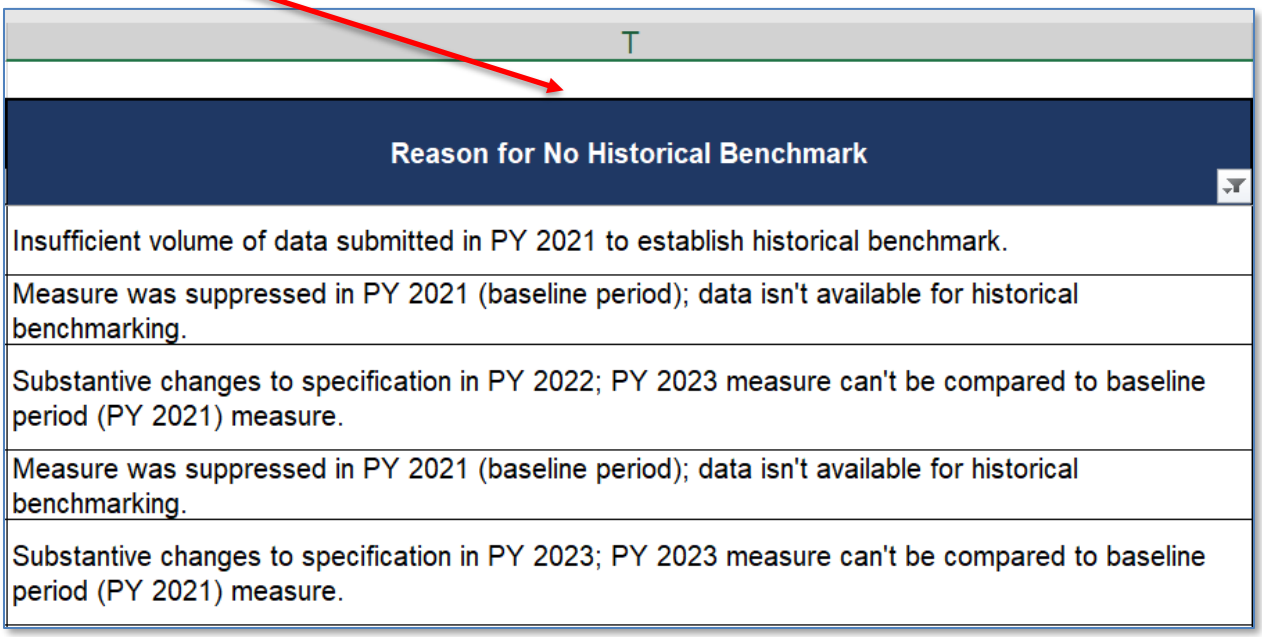
- This includes individual clinicians and groups that are opt-in eligible and elect to opt-in to MIPS participation.
- Voluntary submissions are excluded from benchmark data.

G
Measure has a Benchmark
N

Measures/collection types without **historical** benchmarks display "N" (for "NO") in the "Measure has a Benchmark" column (Column G).

Original Posting: 1/25/2023

Column T in the benchmark file indicates why there's no historical benchmark for a measure/collection type.



T
Reason for No Historical Benchmark
Insufficient volume of data submitted in PY 2021 to establish historical benchmark.
Measure was suppressed in PY 2021 (baseline period); data isn't available for historical benchmarking.
Substantive changes to specification in PY 2022; PY 2023 measure can't be compared to baseline period (PY 2021) measure.
Measure was suppressed in PY 2021 (baseline period); data isn't available for historical benchmarking.
Substantive changes to specification in PY 2023; PY 2023 measure can't be compared to baseline period (PY 2021) measure.

NEW in 2023: If no historical benchmark exists and no performance period benchmark can be calculated, then the measure will receive 0 points.

EXCEPTIONS:

- We've established a policy for new quality measures; please note that these policies don't apply to administrative claims-based measures.
    - New measures in their 1st year in the program are subject to a 7-point scoring floor provided data completeness requirements are met.
    - New measures in their 2nd year in the program are subject to a 5-point scoring floor provided data completeness requirements are met.
- Small practices will continue to receive 3 points for measures without a benchmark, even if data completeness and case minimum requirements aren't met.

Original Posting: 1/25/2023

Are All Topped Out Measures Capped At 7 Points?

No. A measure is capped at 7 points when it is topped out through the same collection type for 2 (or more) consecutive years. The 7-point cap is applied in the 2nd year the measure is identified as topped out.

A measure may be topped out without being capped at 7 points. A "Yes" in the **Seven Point Cap** column (column S) of the benchmark file indicates the measure is capped at 7 points.

Example 1. Measure ID 127, **Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan** (MIPS CQM)

Topped Out	Seven Point Cap
Yes	No

Even though it's topped out, it's not capped at 7 points because it wasn't topped out last year. (It's not in its 2nd consecutive year of being topped out.)

A maximum of 10 achievement points is available for the measure.

Example 2. Measure ID 130, **Documentation of Current Medications in the Medical Record** (all collection types)

This measure is has been topped out for at least 2 consecutive years. A maximum of 7 achievement points is available for the measure, even if your performance rate is found in Deciles 7 – 10.

Topped Out	Seven Point Cap
Yes	Yes

Did you know?

The benchmark file displays the range of performance rates associated with Deciles 7 – 10, even though scoring is capped at 7 points.

Original Posting: 1/25/2023

How Do Benchmarks Work for Multi-Performance Rate Measures?

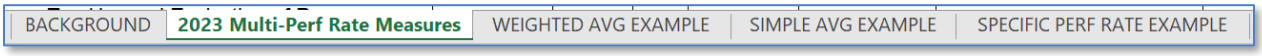
Several MIPS quality measures and QCDR measures require the collection and submission of data for multiple populations. Therefore, multiple performance rates can be associated with a single measure.

- Historical benchmarks for multi-performance rate measures are created based on an "overall performance rate" (based on a weighted average, simple average, or CMS-specified performance rate).
- When you are scored on a multi-performance rate measure, we'll compare the "overall performance rate" of your submitted measure to the measure's benchmark which is also based on the "overall performance rate."

The **2023 Multi-Performance Rate Measure file** identifies the method used to determine the "overall performance rate" for each multi-performance rate measure.

- It ISN'T intended to specify an additional performance rate that must be submitted. Measures should be submitted according to their specification.
- Only multi-performance rate QCDR measures allow for the submission of an "overall performance rate."
- We'll use the "overall performance rate" to determine whether data completeness and case minimum criteria have been met for the measure.

This file also provides an example for each of the 3 methods for determining the overall performance rate. (Click the tabs at the bottom of the file.)



Where Can I Find Performance Period Benchmarks?

We'll publish 2023 performance period benchmarks on the QPP Resource Library once they're available in summer 2024.

Original Posting: 1/25/2023

Where You Can Go for Help

- Contact the Quality Payment Program Service Center at 1-866-288-8292, Monday through Friday, 8 a.m. – 8 p.m. ET or by e-mail at: QPP@cms.hhs.gov.
    - Customers who are hearing impaired can dial 711 to be connected to a TRS Communications Assistant.
- Visit the Quality Payment Program website for other help and support information, to learn more about MIPS, and to check out the resources available in the Quality Payment Program Resource Library.

Version History

If we need to update this document, changes will be identified here.

Date	Change Description
01/25/2023	Original posting.

Original Posting: 1/25/2023