



ACUMEN

**Physician Compare Quality Measurement
Technical Expert Panel (TEP) Summary Report**

December 2020

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010

TABLE OF CONTENTS

- 1 Introduction..... 3**
- 2 About the TEP..... 4**
- 3 Quality Payment Program Background 6**
 - 3.1 The Quality Payment Program 6
 - 3.1.1 The Merit-Based Incentive Payment System..... 6
 - 3.1.2 Alternative Payment Models 7
 - 3.2 PY 2017 Performance Data Release..... 7
 - 3.3 Public Reporting Plan 8
 - 3.3.1 Quality Payment Program Provisions..... 8
 - 3.3.2 Data Analysis Plan..... 9
 - 3.3.3 Feedback on Public Reporting Recommendations 13
- 4 PY 2018 QPP Public Reporting Considerations 14**
 - 4.1 Expansion of Reporter Population – Voluntary Reporters 14
 - 4.2 Expansion of Star Ratings and Collection Types 15
 - 4.3 Expansion of Eligible Performance Information..... 17
 - 4.3.1 MIPS Quality Outcome Measures 17
 - 4.3.2 MIPS Promoting Interoperability Measures and Attestations 18
 - 4.3.3 MIPS Improvement Activities 19
- 5 Final Performance Year 2018 Public Reporting Plan 20**
 - 5.1 Measures Recommended for Public Reporting 20
 - 5.1.1 PI Category 21
 - 5.1.2 IA Category..... 21
 - 5.1.3 MIPS Final and Performance Category Scores 21
 - 5.1.4 APM Participation 21
 - 5.1.5 ACOs 22

LIST OF TABLES AND FIGURES

- Table 1. TEP Members 4
- Table 2. PY 2017 Performance Information Reported on Physician Compare 8
- Table 3. Comparison of Star Rated Measures, PY 2017 and PY 2018 15
- Table 4. Quality Outcome Measure Recommendations 18
- Table 5. Clinician APM Participation and Public Reporting of data on Physician Compare 22

1 INTRODUCTION

As CMS's primary platform for patients and caregivers to review Medicare-accepting clinicians, the Physician Compare¹ website publishes information about eligible groups and clinicians that is carefully selected by CMS on a systematic iterative basis. In 2020, performance information from performance year (PY) 2018 of the Quality Payment Program (QPP) will be published on Physician Compare, which will continue incorporating the requirements outlined in the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) and the calendar year (CY) 2018 QPP Final Rule (82 FR 53819 through 53832). The CY 2018 QPP Final Rule establishes that, in addition to other types of QPP indicators and affiliations, all performance information submitted through the Merit-Based Incentive Payment System (MIPS) are available for public reporting (82 FR 53819 through 53832). In addition, the CY 2019 QPP Final Rule expanded the use of the Achievable Benchmark of Care (ABC™) methodology for establishing QCDR measure benchmarks beginning with the 2018 performance year (83 FR 59915). MACRA builds upon the Physician Compare requirements outlined in Section 10331 of the Affordable Care Act. To meet the rigorous public reporting standards established in Section 10331 (a)(1) of the Affordable Care Act and through rulemaking, all QPP performance information selected for public reporting must be accurate, valid, reliable, and comparable across available collection types, and any information included on clinician, group, and/or Accountable Care Organization (ACO) profile pages must also resonate with patients and caregivers, as shown through user testing. CMS has contracted the Physician Compare support team to ensure the PY 2018 performance information aligns with statutory and regulatory obligations for public reporting. As part of that process, the team convened two meetings in 2019 with the Physician Compare Technical Expert Panel (TEP) to obtain feedback on the intended approach to public reporting for PY 2018.

The remainder of this report summarizes the discussions and conclusions from both the May and November 2019 TEP meetings. Section 2 introduces the Physician Compare TEP. Section 3 reviews the QPP, the PY 2017 performance information release, and the public reporting standards set forth for Physician Compare. Section 4 describes topics addressed during the TEP meetings. Section 5 outlines the final PY 2018 public reporting plan.

¹ Physician Compare is defined as Physician Compare and/or successor website.

2 ABOUT THE TEP

The Physician Compare support team consults with the Physician Compare Technical Expert Panel (TEP) for guidance on how to choose and display performance information on Physician Compare in a way that accurately and robustly reflects clinical performance and supports actionable comparisons. The TEP consists of clinicians, purchasers, and other experts with a broad range of experience in publicly reporting performance measures, improving health care quality, and developing and testing quality measures (Table 1). The Physician Compare support team convened the TEP on May 15, 2019 to discuss the data availability and analysis plan, policy considerations, and concept testing for publicly reporting PY 2018 MIPS performance information. TEP members reconvened on November 6, 2019 to review considerations surrounding policy decisions and public reporting recommendations for PY 2018. Table 1 lists TEP participants and which TEP meeting they attended.

Table 1. TEP Members

TEP Member	Position(s), Organization	Location	TEP Attended
A.J. Yates, MD	Associate Professor, Department of Orthopedic Surgery/University of Pittsburgh School of Medicine	Pittsburgh, PA	May & November
Dale Shaller, MPA (TEP Chair)	Principal, Shaller Consulting Group	Stillwater, MN	November
Eric Holmboe, MD	Internist, Senior Vice President, Milestones Development and Evaluation of the Accreditation Council for Graduate Medical Education (ACGME)	Philadelphia, PA	-
Gregory Dehmer, MD	Professor of Medicine at the Texas A&M University College of Medicine and Director of the Cardiology Division at the Scott & White Clinic	Temple, TX	May & November
Jeffrey P. Jacobs, MD	Director of ECMO Program at All Children's Hospital, Professor of Cardiac Surgery (PAR) in the Division of Cardiac Surgery of the Department of Surgery at Johns Hopkins University, Surgical Director of the Heart Transplantation Program at All Children's Hospital, and Clinical Professor in the Division of Thoracic/Cardiovascular Surgery at University of South Florida College of Medicine.	St. Petersburg, FL	May & November
Michael Mihlbauer, MS	Practice Administrator, Anesthesiology Associates of Wisconsin	Milwaukee, WI	-
Robert Krughoff, JD	Founder and President, Center for the Study of Services/Consumers' Checkbook	Washington, DC	November

TEP Member	Position(s),Organization	Location	TEP Attended
Sara Scholle, DrPH	Assistant Vice President, Research & Analysis/National Committee for Quality Assurance ²	Washington, DC	November
Sherrie Kaplan, PhD, MSPH, MPH	Professor of Medicine and Assistant Vice Chancellor, Healthcare Evaluation and Measurement Executive Co-Director, Health Policy Research Institute School of Medicine/ University of California, Irvine	Irvine, CA	May & November
Ted von Glahn, MS	Consultant	San Francisco, CA	May & November
Thomas Smith, MD, MS	Medical Director, Division of Managed Care, NYS Office of Mental Health/New York State Psychiatric Institute	New York, NY	-

² The National Committee for Quality Assurance joined the Physician Compare project team in August, 2019, though Dr. Scholle was not directly involved in that work.

3 QUALITY PAYMENT PROGRAM BACKGROUND

CMS publicly reports Quality Payment Program (QPP) performance information on Physician Compare clinician, group, and Accountable Care Organization (ACO) profile pages and in the downloadable files. All performance information must meet the established public reporting standards to be publicly reported (§414.1395(b)).

The remainder of this section discusses considerations pertaining to the publication of performance information collected under the QPP. Section 3.1 describes the QPP and the requirements establishing which QPP performance information must be publicly reported. Section 3.2 summarizes the PY 2017 performance information released on Physician Compare in 2019. Section 3.3 outlines the PY 2018 performance information public reporting plan.

3.1 The Quality Payment Program

The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA), an amendment to Title XVII of the Social Security Act, established the Quality Payment Program (QPP) (82 FR 53569). Per MACRA, clinicians can participate in QPP via one of two tracks: (1) the Merit-based Incentive Payment System (MIPS) or (2) an Advanced Alternative Payment Model (APM) (82 FR 53569). The remainder of this section discusses how performance data are submitted and used through these two options.

3.1.1 *The Merit-Based Incentive Payment System*

Clinicians and groups participating in MIPS must each qualify as a MIPS eligible clinician (EC) (82 FR 53578 through 53579); additional clinicians may submit performance information as voluntary reporters. Aiming to reward high value, high quality Medicare clinicians with payment increases, the MIPS consolidates three legacy programs: the Physician Quality Reporting System (PQRS), the EHR Incentive program, and Value-Based Modifier, as specified under Section 101(b) in the MACRA. MIPS performance is evaluated through four performance categories: Quality, Improvement Activities (IA), Promoting Interoperability (PI),³ and Cost.

Under the QPP, certain performance categories are analogous to legacy programs. The MIPS Quality category replaced the legacy PQRS and is represented on Physician Compare as MIPS Quality measures, CAHPS for MIPS measures, and Qualified Clinical Data Registry (QCDR) quality measures. Performance on MIPS Quality measures is displayed as star ratings and percentages⁴ on live-site profile pages and in the downloadable files. The MIPS Promoting Interoperability (PI) performance category replaces the EHR Incentive Program and attestations.

³ CMS renamed Advancing Care Information and the EHR Incentive Programs to Promoting Interoperability in 2018.

⁴ MIPS quality measures are represented as star ratings on clinician and group profile pages, and as percentages on ACO profile pages.

The MIPS Cost category replaces the legacy Value-Based Payment Modifier Program and includes measures of overall spending as well as measures that are focused on episodes of care based on procedures or patient conditions.

The MIPS final score is the sum of the points earned under each performance category with relevant weighting applied; each performance category's assigned weight is multiplied by 100 (82 FR 53778). The weighting of the categories will change as the program matures over the initial performance years. The weighting for PY 2017: Quality (60%), Cost (0%), IA (15%), and PI (25%) and PY 2018: Quality (50%), Cost (10%), IA (15%), and PI (25%) (82 FR 53779) reflects the growth of the MIPS Cost performance category. The weight of these performance categories may be redistributed depending on the measures available to a particular group or individual clinician or extenuating circumstances that occur during the reporting period (82 FR 53779 through 53785) on a case-by-case basis.

3.1.2 Alternative Payment Models

Eligible clinicians can participate in the QPP through Alternative Payment Models (APMs), which can be classified as APMs, Advanced APMs, or MIPS APMs. MIPS APMs are also subject to the APM Scoring Standard under MIPS (82 FR 53899). In order to qualify as an Advanced APM, participants must (1) use certified electronic health record technology (CEHRT), (2) provide payment for covered professional services based on quality measures comparable to MIPS Quality measures, and (3) either be a Medical Home Model expanded under CMS Innovation Center authority or bear a significant financial risk (82 FR 53834). The PY 2018 performance information submitted by non-ACO Advanced APMs or their Qualified APM Participants (QPs) will not be reported on Physician Compare.

MIPS APM participants include MIPS eligible clinicians who are eligible for the APM scoring standard under MIPS. MIPS APMs must participate in the APM under an agreement with CMS, include one or more MIPS eligible clinicians on a Participation List, and base payments on performance, cost/utilization, and quality measures (82 FR 53899). MIPS APM participants scored under the APM scoring standard will receive a MIPS Final Score and associated MIPS payment adjustment based on the APM entity's combined performance. Per Section 1848(q)(9)(A)(i)(I) that requires all MIPS Final and Performance Category scores are publicly reported, MIPS scores achieved by MIPS APM participants will be made publicly available. PY 2018 MIPS performance information submitted as an individual by clinicians who are not QPs in an Advanced APM may be publicly reported on the Physician Compare.

3.2 PY 2017 Performance Data Release

CMS published the QPP performance information for the first time on Physician Compare in 2019. The PY 2017 performance information publicly reported on Physician

Compare in 2019 spanned both the MIPS Quality and MIPS Advancing Care Information⁵ performance categories; however, only quality measure information was available on the live-site profile pages. Table 2 outlines the number of measures and attestations and display type published on Physician Compare in 2019, by measure and reporting entity (i.e. individual, group, or ACO).

Table 2. PY 2017 Performance Information Reported on Physician Compare

Entity Type	Measure Type	Publication Location	Number of Measures	Display
Quality				
Group	MIPS	Live-Site Profile Pages	12	Star Rating
Group	MIPS	Downloadable	107	Performance Rate Percentage, Benchmark Rate Percentage, and Star Rating
Group	QCDR	Live-Site Profile Pages	6	Performance Rate Percentage
Group	QCDR	Downloadable	7	Performance Rate Percentage
Group	CAHPS for MIPS	Live-Site Profile Pages	8	Mean Performance Percentage
Group	CAHPS for MIPS	Downloadable	8	Mean Performance Percentage
Individual	MIPS	Live-Site Profile Pages	0	N/A
Individual	MIPS	Downloadable	108	Performance Rate Percentage
Individual	QCDR	Live-Site Profile Pages	11	Performance Rate Percentage
Individual	QCDR	Downloadable	13	Performance Rate Percentage
ACO	MIPS	Live-Site Profile Pages	8	Performance Rate Percentage
ACO	CAHPS for MIPS	Live-Site Profile Pages	5	Mean Performance Percentage
Advancing Care Information (Promoting Interoperability)				
Group	Measures	Downloadable	7	Performance Rate Percentage
Group	Attestations	Downloadable	4	Attestation
Individual	Measures	Downloadable	7	Performance Rate Percentage
Individual	Attestations	Downloadable	4	Attestation

3.3 Public Reporting Plan

As Section 3.2 discussed, Physician Compare publicly reported the PY 2017 QPP performance information in 2019. Section 3.3.1 outlines provisions for public reporting as stipulated in the CY 2018 QPP Final Rule. Section 3.3.2 outlines the data analysis plan for assessing that PY 2018 performance information meets public reporting standards.

3.3.1 Quality Payment Program Provisions

Sections 1848(q)(9)(A) and (D) of the Social Security Act require Physician Compare to publicly report (1) MIPS eligible clinicians' final scores, (2) MIPS eligible clinicians'

⁵ Effective with PY 2018, this performance category is now called MIPS Promoting Interoperability.

performance scores for each MIPS category, (3) APM affiliations, and, to the extent feasible (4) the names and performance of APMs. To guide the public reporting of QPP data on Physician Compare, the CY 2018 QPP Final Rule (82 FR 53819 through 53832) outlines the approach for public reporting MIPS and APM data on Physician Compare, as well as any other information required by MACRA. This includes all measures and activities reported under MIPS via all available collection types. CMS has the flexibility to stipulate, via regulation, what data collected under the QPP are published to Physician Compare, where information is published (i.e. live-site profile pages or downloadable files), and how the information is displayed to the public.

The CY 2018 QPP Final Rule continues to expand on the performance information available on the Physician Compare website. CMS will provide MIPS Final and Performance Category Scores for MIPS eligible clinicians for all MIPS performance categories. Performance information from all MIPS categories will be made available, where technically feasible, on clinician, group, and ACO profile pages and/or in the downloadable files. A successful performance indicator for objectives, activities and measures reported under the MIPS PI category⁶ will be included on profile pages, where applicable. MIPS Quality and Cost measures in their first two years of use will not be reported publicly, however, CMS will allow the reporting of MIPS PI and MIPS IA measures and activities in the first year of use. The Achievable Benchmark of Care (ABC™) methodology for benchmarking, which is discussed in Section 3.3.2, will be applied to all categories of MIPS from PY 2018 onward, as technically feasible. The use of the ABC™ methodology for QCDR measures was established in the CY 2019 QPP Final Rule, but is effective for PY 2018 performance information (83 FR 59915). Performance information⁷ for all eligible clinicians and groups who participate in MIPS that meet all public reporting standards may publicly reported, with the exception of those who choose to opt out of public reporting during the data Preview Period.

The Physician Compare team helps to inform public reporting decisions by conducting in-depth analyses on MIPS performance information submitted to CMS to identify performance information that meets public reporting standards. All performance information undergoes statistical testing and website user testing to determine how and where the measures will be publicly reported. Section 3.3.2 outlines the measure analysis approach that helps to inform CMS decision-making.

3.3.2 Data Analysis Plan

The Physician Compare support team conducts a standard set of analyses to identify performance measures that comply with mandated public reporting standards and are available for CMS to publicly report. These standards specify that published performance information

⁶ The successful performance indicator indicates MIPS PI Performance Category score of greater than zero.

⁷ All performance information submitted to CMS under the QPP is eligible for public reporting, regardless of whether it was used for calculating MIPS Final and Performance Category scores.

must be reliable, valid, comparable, and accurate across available performance information collection types. Additionally, concept testing is conducted to ensure that patients and caregivers can use the information published for their decision-making.

ABC™ Benchmarking and Star Rating

The 2016 PFS Final Rule (80 FR 71116 through 71135) established that Physician Compare use the ABC™ benchmark methodology to calculate a 5-star rating benchmark for performance measures. Through discussions at the 2017 TEP meetings,⁸ it was determined that the equal ranges method would be used to produce meaningful and reliable 1- to 4-star rating assignments. A more detailed overview of CMS's star rating and benchmarking methodology can be found on the Physician Compare Initiative Page.⁹

Reliability Tests

Measure reliability refers to the extent to which differences in performance rates for each quality measure are due to actual differences in performance versus variation that arises from measurement error. In order to determine that measure reliability standards are met, reliability is assessed using two methods: (1) the beta binomial test and (2) a split half reliability test. If a measure¹⁰ passes both of these reliability tests, the performance scores for quality measures are considered meaningfully different across reporting entities, rather than due to measurement error. Further, in order to be published as a star rating, the ABC™ benchmark, star rating cut-offs, and star rating assignments must pass further reliability testing.

Reliability of Star Rating Cut-offs

To assess the reliability of the benchmark and star rating cut-offs, the Physician Compare support team tests the stability of star rating cut-offs when presented with changes in the performance rate distribution that could be expected due to chance, given the sample size (e.g. the number of reporters) and the amount of variation in performance across reporting entities. To investigate the production of reliable cut-offs, a bootstrapping analysis is conducted where reporters' performance scores are randomly sampled with replacement until the sample size is equal to the number of reporters for that measure. This process is repeated 500 times for each measure.¹⁰ For each simulated data set, the star rating cut-offs are recalculated and the simulated cut-offs are used to reassign each reporter to a simulated star rating category, using their original performance score.

⁸ [Physician Compare TEP Summary Report, December 2017](#)

⁹ [Physician Compare Benchmark and Star Ratings Fact Sheet](#)

¹⁰ Evaluated at the measure-, measure stratum-, reporting entity-, collection type-level.

Star Rating Assignment Reliability

In addition to producing robust cut-offs, an ideal star rating method should categorize reporters meaningfully, given the precision of the observed performance rates. To ensure that star assignments are not influenced heavily by random error, each reporter's actual performance rate and patient population sizes are used to simulate counterfactual performance rates using a binomial distribution. The ABC™ benchmark and the star rating cut-offs are recalculated for each simulation, and reporters are assigned a simulated star rating category based on their simulated performance rate. After running 500 simulations for a measure/mechanism combination, the frequency of reporters receiving the same and different star ratings across simulations are evaluated.

Reliability Thresholds

For both the star rating cut-off and star rating assignment evaluations, the following thresholds are used to assess reliability: 80% accuracy of assignments across simulations; multi-star shift of less than five percent; and a Fleiss' Kappa of less than or equal to 0.6. Positive results on these tests imply that reporters assigned to different star categories are meaningfully different and can be compared to each other.

Validity

Validity refers to the degree to which a metric measures what it purports to measure. Validity for each quality measure is assessed by evaluating the extent to which observed performance rates on measures are impacted by factors unrelated to true performance, such as characteristics of the reporter's patient population (i.e. case-mix) or selective reporting of the patient population. The validity of measure data is addressed in three ways: (1) evaluation of outcome measure risk adjustment strategies, (2) investigation into the possibility of selective reporting, and (3) review of specifications to ensure they align with clinical best practice guidelines. If outcome measures are not appropriately risk adjusted, measure performance could be influenced by confounding factors, such as patient population characteristics. This would impact the observed performance rates, which would therefore not accurately reflect the true quality of clinical care provided. Therefore, a team of clinical experts reviews the validity of all eligible outcome measure specifications to assess for: (1) scientific acceptability (e.g. consistency with current clinical guidelines and literature); (2) feasibility (e.g. precision of specifications resulting in consistent implementation); (3) usability (e.g. ease of clinician understanding of attribution, risk adjustment, and outcome); (4) importance (e.g. potential for facilitating practice improvement); and (5) measure harmonization (e.g. alignment of goals with currently available quality or cost measures across reporting programs).

Additionally, the Physician Compare team analyzes reporting rates and performances rates for each measure using a Pearson correlation to assess if there is evidence suggesting that

reporters could be selectively reporting data in order to inflate performance scores. Finally, as clinical guidelines and best practices change over time, our team ensures that only measures that align with current best practice guidelines are selected for public reporting.

Comparability

Comparing the performance score distributions from data collected via different collection types (e.g. CMS Web Interface vs. Qualified Registry) for the same measure shows that raw performance data submitted through distinct pathways are not always comparable. Thus, measure analyses and the establishment of benchmarks and star ratings are performed for each measure/collection type¹¹ combination, rather than aggregating data from different mechanisms. Further, the use of benchmarks helps to ensure performance, relative to what is achievable for a given collection type, can be validly compared across collection types.

In previous years, to ensure comparability, CMS opted to select performance information from only one collection type for a given measure for public reporting. When performance information collected via multiple collection types for a single measure met all of the Physician Compare public reporting standards, the collection type that represented the highest number of distinct reporters was selected. However, as part of CMS's phased approach to public reporting, starting with PY 2018, measure performance information from any collection type that meets public reporting standards for publication as star rating on profile pages will be publicly reported. In other words, if multiple measure/collection type¹² combinations for a single measure meet all public reporting standards for public reporting as star ratings on profile pages, all collection types that meet these standards will be made publicly available. If a measure does not meet all public reporting standards for reporting as a star rating on profile pages, the collection type that represents the highest number of distinct reporters will be selected for public reporting as percent performance scores exclusively in the downloadable files.

Accuracy

Measure accuracy refers to the degree to which a measure correctly assesses what it purports to measure. To allow for ample time for measure testing and validity assessments of newly introduced measures, CMS does not publicly report quality or cost measures in their first two years of use (82 FR 53824 through 53826). Additionally, measures with specifications that are found to misalign with best clinical practice are excluded from public reporting, as data submissions for these measures are less likely to accurately reflect true clinical performance.

¹¹ Evaluated at the measure-, measure stratum-, reporting entity-, collection type-level.

¹² Evaluated at the measure-, measure stratum-, reporting entity-, collection type-level.

User Testing

In order to meet public reporting standards, data published on Physician Compare must resonate with users as determined through testing (82 FR 53822 through 53826). The Physician Compare support team tests performance information available for public reporting using plain language to facilitate user understanding. All measures and activities are tested using plain language titles and descriptions before they are selected for public reporting. Additionally, other QPP concepts for potential inclusion on group and/or clinician profile pages such as APM affiliations and the MIPS Promoting Interoperability successful reporter indicator are tested along with website design layouts for their inclusion.

3.3.3 Feedback on Public Reporting Recommendations

Results from the above analyses were used to curate a list of measures and activities recommended for public reporting for PY 2018. This list, along with other considerations in regards to how and where to publish measure data, are presented to the TEP in order to garner expert feedback that CMS may utilize when rendering final public reporting decisions. The following sections of this report detail the discussions and TEP recommendations made in regards to publicly reporting PY 2018 performance information.

4 PY 2018 QPP PUBLIC REPORTING CONSIDERATIONS

The Physician Compare TEP convened in 2019 to discuss several options for implementing public reporting of 2018 QPP performance information. The remainder of this section provides a summary of the content covered during the two 2019 TEP meetings, and where applicable, a summary of TEP member feedback. Section 4.1 describes the discussion of expanding the population of reporters with performance information eligible for public reporting to voluntary reporters. Section 4.2 outlines the expansion of star ratings to both clinicians and groups, multiple MIPS performance categories, and multiple collection types per measure. Section 4.3 details the expansion of eligible performance information available for public reporting with PY 2018, including MIPS Quality outcome measures, the MIPS Promoting Interoperability Objectives and Measures set, and the attestations collected under the MIPS Improvement Activities performance category.

4.1 Expansion of Reporter Population – Voluntary Reporters

As finalized through rule making and discussed in Section 3.3.1, performance information voluntarily submitted to CMS by clinicians and groups that are not eligible for a MIPS payment adjustment is available for public reporting. Voluntary reporters are able to opt out of having their performance information publicly reported during the Physician Compare Preview Period through an opt-out mechanism implemented in the Preview Portal (82 FR 53830). Voluntary reporter and opt-out eligibility status is determined at the individual clinician-practice level using the unique combination of the national provider identification (NPI) number and tax identification number (TIN) for each clinician.¹³

It is possible that some clinicians that are employed in multiple practices are considered MIPS eligible under at least one TIN/NPI combination, but not others. One TEP member asked about opt-out for voluntary data submission at the group versus the individual clinician-level. The opt-out mechanism in the Preview Portal accounts for disparate eligibility between different groups and individuals, and across an individual clinician's various TIN/NPI combinations. Given that a clinician may report as an individual or as part of a group, a TEP member questioned the impact of the voluntary report policy on measures specifications, however, it was determined by the team that the measures specifications would not be affected.

Overall, including voluntary reporter submissions in PY 2018 analyses resulted in more measures passing public reporting standards. Including the voluntary reporters improved

¹³ In this first year of voluntary reporter public reporting eligibility, MIPS performance information submissions by Qualifying Participants (QPs) in Advanced Alternative Payment Models (APMs) will continue to be ineligible for public reporting, but clinicians participating in MIPS APMs will be able to opt out of public reporting for the performance information submitted at the individual clinician-level under the same TIN/NPI used to participate in the MIPS APM.

measure reliability and precision overall. For the vast majority of cases, star rating cut-offs did not change when including submissions by voluntary reporters.

4.2 Expansion of Star Ratings and Collection Types

Since the initial release of 13 group-level PY 2016 PQRS quality measures as star ratings on Physician Compare in 2017, CMS has implemented a phased approach to publicly reporting performance information on Physician Compare profile pages as star ratings. As part of this phased approach, CMS is expanding the set of performance metrics eligible for display as star ratings and the entities that are eligible for star ratings on their profile page. Beginning with the PY 2018 performance information release, all MIPS Quality¹⁴, QCDR, and MIPS PI performance rate measures published on clinician and group profile pages will be represented as star ratings, given they meet all public reporting requirements. Further, the PY 2018 performance information release also marks the first year that performance information from multiple collection types (e.g. EHR, CMS Web Interface, Qualified Registry) per measure are eligible for public reporting, given all entities who reported through the collection type can be reliably categorized into a star rating category.

To date, CMS has released a small subset of PY 2016 PQRS quality measures and PY 2017 MIPS Quality measures as star ratings on group profile pages. However, with the PY 2018 MIPS performance information release, CMS plans to publicly report all clinician- and group-level PY 2018 MIPS Quality, QCDR, and MIPS PI performance rate measures that meet the public reporting standards outlined in Section 3.3. Table 3 below presents a comparison in the number of performance rate measures selected for public reporting as star ratings in PY 2017 and PY 2018. Approximately 20,000 groups and 115,000 individual clinicians reported at least one of the measures recommended for public reporting as star ratings and may have at least one star rating on their profile page.

Table 3. Comparison of Star Rated Measures, PY 2017 and PY 2018

Measure Type	Groups		Individual Clinicians	
	PY 2017	PY 2018	PY 2017	PY 2018
MIPS Quality	12	84	0	77
QCDR	0	9	0	9
MIPS Promoting Interoperability ¹⁵	0	13	0	13

Raw performance rate data from separate collection types for the same measure are not always comparable for public reporting purposes. Rather than selecting a single collection type

¹⁴ The CAHPS for MIPS summary survey score measures will not be publicly reported as star ratings, but will continue to be represented as top-box percent performance scores on group profile pages.

¹⁵ In PY 2017, this category was called MIPS Advancing Care Information

per measure for public reporting, starting with the PY 2018 performance information release, CMS will publicly report performance information from multiple collection types per measure, given they meet all public reporting standards and reporters can be reliably categorized into star ratings. Performance benchmarks and star ratings are established independently for each reporting entity/measure/collection type combination, which results in a star rating that represents reporter performance relative to their peers subject to the same measure specification and reporting protocols. Standardizing performance using star ratings, which are well understood by patients and caregivers according to our user testing, allows users to accurately compare relative performance across collection types. Expanding star ratings to include all collection types impacts approximately 2,100 groups and 12,000 individual clinicians who would not otherwise be eligible for any star ratings if only a single collection type was chosen per PY 2018 measure.

Because not all measures meet the public reporting standards required for publication on profile pages as star ratings, the approach used to select collection types for public reporting is as follows:

- If all collection types for a reporting entity/measure meet public reporting standards for display as star ratings, then all collection types are recommended for public reporting;
- If only some of the available collection types for a reporting entity/measure meet public reporting standards for display as star ratings, then only the collection types that meet these standards are recommended for public reporting;
- If none of the available collection types for a reporting entity/measure meet public reporting standards for display as star ratings, then the most frequently reported collection type that meets public reporting standards for display as a percent performance score will be selected for public reporting exclusively in the downloadable files.

The TEP did not express concerns related to the expansion of star ratings from a small subset of group-level measures to all clinician- and group-level performance rate measures that pass all public reporting standards for publication as star ratings. The TEP members did weigh the pros and cons of establishing different benchmarks for each collection type per measure. For example, one TEP member noted that clinicians or groups may be incentivized to report their performance information for a measure through a collection type with a lower 5-star benchmark to maximize their star rating. However, another TEP member noted that the hurdles of changing reporting practices may outweigh this possible gain. Another TEP member suggested incorporating collection type validating considerations into rule making, such as stating that if the collection type diverges considerably from other distributions it would not be considered for public reporting.

4.3 Expansion of Eligible Performance Information

The performance information publicly reported on Physician Compare has evolved over the last several years with the shift to star ratings and the publication of data from the first QPP performance year. The type of performance information reported on Physician Compare has expanded with the introduction of QPP, and now includes measure and attestation data from the various MIPS performance categories, MIPS final and performance category scores, and also includes APM affiliations. As mentioned in Section 4.2, the number of measures eligible for public reporting as star ratings will increase significantly with the PY 2018 performance information release. In addition, with the PY 2018 performance information release, CMS will continue to expand the MIPS Quality, Promoting Interoperability (PI) measures and attestations, and Improvement Activities (IA) publicly reported on Physician Compare. While the 2018 MIPS Cost performance category score will be publicly reported, the MIPS Cost measures do not meet public reporting standards and will not be publicly reported.

The remaining sections cover the measure types and activities that are newly available with the PY 2018 performance information release, including MIPS Quality outcome measures, the MIPS Promoting Interoperability Objectives and Measures, and the MIPS Improvement Activities attestations.

4.3.1 MIPS Quality Outcome Measures

Quality outcome measure performance is highly valuable to CMS, Medicare beneficiaries, and their caregivers. CMS prioritizes the reporting of outcome measures in MIPS by requiring at least one outcome measure to be reported in the quality category and offering bonus points for the reporting of additional outcome measures. Accordingly, CMS plans to publicly report the MIPS Quality outcome measures that meet all public reporting standards, as outlined in Section 3.3.2. A small portion of the available PY 2018 outcome measures passed the minimum eligibility criteria (e.g. must have at least 20 reporters), those that passed this criteria were further evaluated for clinical and face validity prior to considering which ones should be eligible for public reporting. Specification validity was assessed by a team of clinicians and clinical experts who evaluated the scientific acceptability, feasibility, usability, importance, and measure harmonization.

Table 4 below displays the PY 2018 MIPS Quality outcome measures that will be publicly reported for individual clinicians and/or groups. The table shows the measure number, the measure title, and whether the measure was selected for public reporting in a downloadable format only (“DL”) or both on live-site profile pages and the downloadable files (“LS”).

Table 4. Quality Outcome Measure Recommendations

Msr. #	Msr. Title	Individual	Group
191	Cataracts: 20/40 or Better Visual Acuity within 90 Days Following Cataract Surgery	DL	LS
192	Cataracts: Complications within 30 Days Following Cataract Surgery Requiring Additional Surgical Procedures	-	DL
303	Cataracts: Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery	LS	-
342	Pain Brought Under Control Within 48 Hours	DL	-
370	Depression Remission at Twelve Months	-	LS
398*	Optimal Asthma Control	DL	-
410	Psoriasis: Clinical Response to Oral Systemic or Biologic Medications	DL	LS
435	Quality of Life Assessment For Patients With Primary Headache Disorders	LS	-

*Multi-strata measure

The TEP members focused their questions on the methodology used to evaluate the clinical face validity of the outcome measure specifications. For example, one TEP member asked if a meaningful improvement in a measure represented by a small performance change would be captured and distinguishable across reporters. In follow-up, another TEP member mentioned that when evaluating narrow outcomes, reliability and validity may be lost. The reliability tests the extent to which differences in performance rates for each quality measure are due to actual differences in performance versus variation that arises from measurement error, and only measures that have sufficient reliability are selected for public reporting. One TEP member asked for further clarification about the evaluation of measure harmonization in the clinical face validity assessment process. The aim in evaluating measure harmonization is to evaluate the extent to which measures work together with other measures to create an improved assessment of a clinician’s practice and treatment. The TEP members did not recommend any additional evaluation criteria for quality outcome measures at this time.

4.3.2 MIPS Promoting Interoperability Measures and Attestations

The MIPS PI performance category aims to assess eligible clinician and group commitment to patient engagement and electronic exchange of information using certified electronic health record (EHR) technology (CEHRT). During the first two years of the MIPS program (i.e. PY 2017 and PY 2018), MIPS participants could select from two sets of MIPS Promoting Interoperability objectives and measures sets: the PI Transition Objectives and Measures (the “transition” set) or the PI Objectives and Measures (the “non-transition” set). The non-transition set is only available to reporters with CEHRT certified to the 2015 edition or more recent, whereas the transition set is available to reporters with CERHT certified to the 2014 or 2015 edition. The non-transition set is a more expansive measure set than the transition set,

which was used to facilitate the move from the 2014 to 2015 CEHRT data requirements. In PY 2017, only the transition set was publicly reported, but starting with PY 2018, both sets are available for public reporting. Starting with PY 2019, only the non-transition set will be available under the MIPS program. Both the transition and non-transition measure sets include the e-Prescribing, Patient-Specific Education, Provide Patient Access, Secure Messaging, and View, Download, and Transmit (VDT) measures.

To evaluate the comparability of submissions for the analogous measures across the two sets, we reviewed the specifications, assessed performance distributions for extreme differences, and applied our star rating methodology in two separate methods. Benchmarks and star ratings were evaluated by (1) combining submissions for the analogous measures and (2) rating the transition and non-transition submissions for analogous measures separately. The evaluation demonstrated that 97% of group star ratings and 99% of individual star ratings were identical under both methods. To ensure comparability and align with the approach used for other submissions, benchmarks and star ratings will be calculated separately for each reporting entity/measure/collection type. A single set of plain language titles and descriptions will be used to describe analogous measures on live-site profile pages due to the similarity in concepts and specifications. The TEP members did not raise any concerns with this approach for publicly reporting the MIPS PI transition and non-transition measures and objectives.

4.3.3 MIPS Improvement Activities

The MIPS IA category is newly eligible for public reporting with PY 2018; it was “new” under PY 2017 of the QPP, and therefore not eligible for public reporting at that time (82 FR 53826). This category represents clinicians’ and groups’ participation in activities that improve clinical practice. To meet the category requirements, reporters must attest to at least two to four activities. Reporters can choose from a list of over 100 attestations that represent dedication to practice improvement in realms such as care coordination, emergency response preparedness, and beneficiary engagement. On average, clinicians and groups report between three to four attestations; a small portion of reporters attest to 10 or more activities. In concept testing, we found publicly reporting a long list of attestations on group and clinician profile pages does not resonate well with consumers. Therefore, CMS has elected to limit the number of attestations per page to no more than 10. For reporters who attest to more than 10 activities, CMS will select the 10 most highly reported attestations by the same reporter type for display on profile pages; the downloadable files will include all attestations reported by each clinician and group. When considering future display and functionality options, one TEP member discussed the value of allowing the web user to drive the display to cater to individual preferences. Additionally, other TEP members suggested allowing searching for the IA itself to see a list of clinicians that were reporting on the activity.

5 FINAL PERFORMANCE YEAR 2018 PUBLIC REPORTING PLAN

Based on statistical testing, stakeholder outreach, concept testing, and discussions from the May and November 2019 TEP meetings, CMS has decided to proceed with the following:

5.1 Measures Recommended for Public Reporting

The following information will be eligible for publication on live-site profile pages and in the downloadable files:

- Group- and ACO-level CAHPS for MIPS displayed as a percentage
- Group- and individual-level MIPS Quality measures displayed as a star rating
- ACO-level MIPS Quality measures displayed as a percentage
- Group- and individual-level QCDR measures displayed as a star rating
- Group- and individual-level MIPS PI measures displayed as a star rating
- Group- and individual-level MIPS PI attestations as a checkmark
- Group- and individual-level MIPS IA attestations as a checkmark
- PI successful reporter indicator as a checkmark

In addition to all items listed above, the following data will be available exclusively in the downloadable files:

- MIPS Final score
- MIPS Quality Performance Category score
- MIPS PI Performance Category score
- MIPS IA Performance Category score
- MIPS Cost Performance Category score
- Clinician utilization data

In PY 2018 all collection types will be available for reporting including: Electronic Health Record, QCDR, Qualified Registry, CMS Web Interface, Web Attestation, and Claims. New cost and quality measures will not be published on Physician Compare for PY 2018. Cost measures will not be publicly reported for PY 2018, as they do not meet public reporting standards. Clinicians and groups who submitted MIPS performance information but were not MIPS eligible during PY 2018 may have performance information publicly reported on Physician Compare unless they choose to opt out of public reporting during the 60 day preview period. CMS has ultimately recommended suppressing MIPS measures 122, 137, 138, 224, 236, 317, 343, 373, 378 and WCHQ10 from public reporting due to substantive specifications changes applied during PY 2018.

5.1.1 PI Category

PI performance measures will be reported as star ratings, and attestations will be reported as checkmarks for PY 2018. In addition, an indicator will be included on profile pages for clinicians and groups that successfully submitted 2018 MIPS PI performance data (i.e. had a MIPS PI Performance Category score of greater than 0 at the individual clinician or group level, respectively), which was found to resonate well with website users during concept testing.

5.1.2 IA Category

All IAs available in PY 2018 will be reported as indicated by checkmarks for attestations. A maximum of 10 attestations per profile page will be reported according to user preference. For reporters with more than 10 attestations, the 10 most highly reported attestations by entity will be selected for public reporting on their profile pages.

5.1.3 MIPS Final and Performance Category Scores

Because this information does not resonate with Medicare beneficiaries and their caregivers, MIPS Final and Performance Category scores will be published in the downloadable files at the TIN/NPI level. If an individual has multiple MIPS composite scores at the TIN/NPI level, CMS will publicly report the highest final score and respective category scores.

5.1.4 APM Participation

Clinicians or groups will have an indicator of APM participation on their profile pages, if they participated in the following APMs:

- Accountable Health Communities Model
- Bundled Payments for Care Improvement (BPCI) Advanced, Model 2, Model 3, and Model 4
- Comprehensive Joint Replacement (CJR)
- Comprehensive End-Stage Renal Disease (ESRD) Care (CEC)
- Comprehensive Primary Care Plus (CPC+)
- Frontier Community Health Integration Project Demonstration
- Initiative to Reduce Avoidable Hospitalization Among Nursing Facility Residents: Phase 2
- Maryland All Payer Hospital Model
- Medicare Shares Savings Program Accountable Care Organizations (ACOs)
- Million Hearts: Cardiovascular Disease Risk Reduction
- Next Generation ACO Model
- Oncology Care Model
- Transforming Clinical Practice Initiative.

Individual clinician-level data may be published on Physician Compare, depending on the type of APM participation (Table 5).

Table 5. Clinician APM Participation and Public Reporting of data on Physician Compare

APM Participation	Physician Compare Data Release Information
Qualifying APM Participants in Advanced APMs	MIPS performance information submitted by a Qualifying APM Participant (QP) in an Advanced APM as an individual will not be publicly reported on the clinician’s profile page.
Clinicians in MIPS APMs	MIPS performance information submitted by an eligible clinician at the individual clinician level with a TIN/NPI in a MIPS APM may be available for public reporting on their clinician profile page and are eligible for opt-out during the Physician Compare Preview Period.
Clinicians in All Other APM Types	MIPS performance information submitted by an eligible clinician at the individual clinician level in APMs that are neither an Advanced APM nor a MIPS APM may be publicly reported on their clinician profile page.

5.1.5 ACOs

Next Generation or Medicare Shared Savings Program (Shared Savings Program) ACOs will have ACO profile pages with performance scores on Physician Compare. ACO quality measure and CAHPS for ACO performance rates will be publicly reported on ACO live-site profile pages.