



ACUMEN

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

December 2017 (Edited, May 2018)

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

[This page is intentionally left blank.]

TABLE OF CONTENTS

1	Introduction	5
2	About the TEP	6
3	Overview of 2016 Public Reporting Plan	8
3.1	Evolution of Physician Compare Measure Publication	8
3.2	Quality Measure Benchmark Selection Process	9
4	Star Rating Methodology Development	10
4.1	ABC TM Benchmark Calculation	10
4.2	Star Rating Assignment with Candidate Methods	10
4.3	Star Rating Reliability Analyses.....	12
4.3.1	Cut-off Robustness	13
4.3.2	Performance Rate Precision.....	15
4.4	Removal of Major Outliers	17
4.5	Comparison of Candidate Star Rating Methods	19
4.5.1	TEP Input on the Decision between Candidate Methods	19
4.6	Star Rating Methodology Decision.....	20
5	PY 2016 Measure Selection Process	21
5.1	Eligible Measures Data.....	21
5.1.1	Excluded Measures	21
5.1.2	Excluded Reporters.....	22
5.2	Meeting Public Reporting Standards	22
5.2.1	Comparability	22
5.2.2	Reliability.....	23
5.2.3	Validity	24
5.3	Selected Measures.....	26

LIST OF TABLES AND FIGURES

Table 1.	TEP Members	6
Table 2.	Number of Measures Reported on Physician Compare Over Time	8
Figure 1.	PQRS 47 – Advanced Care Plan Group Registry Performance Rate Distribution	11
Figure 2.	Star Rating Distributions – PQRS 47 – Advanced Care Plan Group Registry Data.....	12
Figure 3.	Star Rating Cut-off Reliability Contingency Table – PQRS 47 – Advanced Care Plan Group Registry Data	15
Figure 4.	Star Rating Cut-off Reliability Contingency Table – PQRS 370 – Depression Remission at Twelve Months Group Web Interface Data	15
Figure 5.	Star Rating Assignment Reliability Contingency Table – PQRS 110 – Influenza Immunization Group Web Interface Data.....	16
Figure 6.	Performance Rate Precision Reliability Contingency Table – PQRS 334 – Appropriate Use of CT Scan for Chronic Sinusitis Group Registry Data	17

Figure 7. PQRS 324 – Appropriate Use of Cardiac Stress Imaging Group Registry Raw
Performance Score Distributions 18

Figure 8. Star Rating Distributions – PQRS 324 – Appropriate Use of Cardiac Stress Imaging
with and without major outliers removed 18

1 INTRODUCTION

Physician Compare draws its operating authority from section 10331(a)(1) of the Affordable Care Act. Under this authority, the Centers for Medicare & Medicaid Services (CMS) initiated a phased approach to public reporting on the website. Sections 1848(q)(9)(A) and (D) of the Medicare Access and CHIP Reauthorization Act (MACRA) facilitate the continuation of this phased approach to public reporting. This legislation and regulation together require that data reported on Physician Compare meet established public reporting standards. These require that data reported through Physician Compare are accurate, valid, reliable, and comparable across the available submission mechanisms. Further, any measures published to public-facing profile pages must resonate with, and be accurately interpreted, by website users as determined through user testing. CMS has contracted with the Physician Compare support team, who work directly with the Physician Compare Quality Measurement Technical Expert Panel (TEP) on a regular basis, to ensure the information published on Physician Compare meets these public reporting standards.

The remainder of this report summarizes the discussions and conclusions from these meetings. Section 2 introduces the Physician Compare TEP. Section 3 provides an overview of how Physician Compare expanded over time and how the Physician Compare quality measure benchmark was selected. Section 4 describes the steps taken to arrive at a Physician Compare star rating methodology. Section 5 walks through the process for selecting the quality measures to publish on Physician Compare in late 2017.

2 ABOUT THE TEP

The Physician Compare support team consults with the Physician Compare Quality Measurement Technical Expert Panel (TEP) for guidance on how to choose and display quality metrics in a way that accurately and robustly reflects clinical performance. The TEP consists of representatives of the patient/caregiver dynamic, purchasers, and technical experts with a broad range of experience in publicly reporting performance measures, improving health care quality, and developing and testing quality measures. The Physician Compare support team convened the TEP on February 27, 2017 and June 7, 2017 to discuss the star rating methodology targeted for publicly reporting program year (PY) 2016 quality measures. The Physician Compare support team and the TEP reconvened on August 28, 2017 to discuss the selection of the Physician Quality Reporting System (PQRS) and the non-PQRS Qualified Clinical Data Registry (QCDR) measures to publish to the Physician Compare website in late 2017. Table 1 lists the 11 individuals who comprise the TEP.

Table 2. TEP Members

TEP Member	Position(s), Organization	Location
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
Ted von Glahn, MS	Consultant	San Francisco, CA
Eric Holmboe, MD	Internist, Senior Vice President, Milestones Development and Evaluation of the Accreditation Council for Graduate Medical Education (ACGME)	Philadelphia, PA
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
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
Robert Krughoff, JD	Founder and President, Center for the Study of Services/Consumers' Checkbook	Washington, DC
Michael Mihlbauer, MS	Practice Administrator, Anesthesiology Associates of Wisconsin	Milwaukee, WI
Sara Scholle, DrPH	Assistant Vice President, Research & Analysis/National Committee for Quality Assurance	Washington, DC

TEP Member	Position(s), Organization	Location
Dale Shaller, MPA (TEP Chair)	Principal, Shaller Consulting Group	Stillwater, MN
Thomas Smith, MD, MS	Medical Director, Division of Managed Care, NYS Office of Mental Health/New York State Psychiatric Institute	New York, NY
A.J. Yates, MD	Associate Professor, Department of Orthopedic Surgery/University of Pittsburgh School of Medicine	Pittsburgh, PA

3 OVERVIEW OF 2016 PUBLIC REPORTING PLAN

All measures available for public reporting are determined through rulemaking. The PY 2016 measures were specifically finalized in the Calendar Year (CY) 2016 Physician Fee Schedule (PFS) final rule (80 FR 71116 through 71135). This continued CMS' phased approach to public reporting on Physician Compare. Consistent with that phased approach, beginning in late 2017, CMS plans to phase in the public reporting of associated 5-star ratings based on a benchmark for publication on Physician Compare. Section 3.1 outlines the phased approach to public reporting implemented for Physician Compare to date. Section 3.2 describes the process for selecting a quality measure benchmark for Physician Compare.

3.1 Evolution of Physician Compare Measure Publication

Beginning in 2014, as indicated by the CY 2012 PFS Final Rule (76 FR 73417 through 73422), CMS began a phased approach to publishing quality measures on the Physician Compare website. Initially, only PQRS measures reported by groups through the Web Interface, and those submitted by Shared Savings Program and Pioneer Accountable Care Organizations (ACOs), were available for publication. Over time, CMS gradually introduced data reported by eligible clinicians, a larger set of quality measures (including Consumer Assessment of Healthcare Providers and Systems [CAHPS] for PQRS patient experience summary survey measures), and a wider range of eligible submission mechanisms (e.g. Medicare Part B Fee-for-Service claims). Table 2 outlines the number of measures published on Physician Compare in late 2014, 2015, and 2016, by submission mechanism.¹

Table 2. Number of Measures Reported on Physician Compare Over Time

Submission Mechanism	PY 2013 Data (Published in 2014)	PY 2014 Data (Published in 2015)	PY 2015 Data (Published in 2016)
Group Web Interface	4	14	14
Group Registry	-	-	98
CAHPS for PQRS	-	8	8
Individual Claims	-	6	44
Individual Registry	-	-	56
Individual QCDR	-	-	16*

*The Physician Compare support team deemed 15 additional non-PQRS QCDR measures sufficient for public reporting on the respective QCDR website, for which a link is displayed on the Physician Compare website.

¹ More information about the Physician Compare public reporting plan can be found at: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/physician-compare-initiative/index.html>

3.2 Quality Measure Benchmark Selection Process

For the first time on Physician Compare, in late 2017, measure- and submission-mechanism-level star ratings based on a benchmark are available for public reporting. As the anchor of the star rating system, benchmarks can provide an accurate and simplified point of comparison across complex data and set an achievable standard for quality performance over time. The remainder of this section outlines the steps taken by the Physician Compare support team to finalize a quality measure benchmark.

Since 2014, CMS and the Physician Compare support team have worked with the stakeholder community and the TEP to define a benchmark and star attribution approach that meets established public reporting standards. The CY 2015 PFS proposed rule (79 CFR 40385 through 40391) included a composite benchmarking methodology for public reporting on Physician Compare that aligned with the methodology used under the Medicare Shared Savings Program. However, stakeholders shared concerns about using a composite benchmark that could be influenced by complex patient populations and specifying a list of measures to use in the composite, especially given that individual reporters often select different sets of measures to submit. In response, the Physician Compare support team conducted extensive stakeholder outreach and then reconvened the TEP to discuss a measure-level benchmark.

With Physician Compare's unique data considerations and stakeholder feedback in mind, the TEP suggested the Achievable Benchmark of Care (ABCTM). Historically well-received by the health care professionals and the entities it measures, the ABCTM benchmark produces an achievable standard of quality and has been shown to lead to improved quality of care.² Citing strong stakeholder support, the finalized CY 2016 PFS final rule established that Physician Compare use the ABCTM benchmark to develop star ratings for individual measures. While the PFS rule establishes the ABCTM benchmark as the method for setting the 5-star cut-off, it does not specify the method for assigning the remaining star categories (e.g. 1-star, 2-stars, 3-stars, or 4-stars). Since the finalization of the ABCTM benchmark, the Physician Compare support team has held formal and informal webinars, Physician Open Door Forums (ODF), discussions with CMS leadership and stakeholders, and TEP meetings to guide the extensive analysis undertaken to create a star rating methodology that incorporates the ABCTM benchmark. Section 4 discusses the conversations with the TEP about developing a method that produces meaningful and reliable star rating assignments.

² Weissman, N. W., Allison, J. J., Kiefe, C. I., Farmer, R. M., Weaver, M. T., Williams, O. D., Child, I. G., Pemberton, J. H., Brown, K. C. and Baker, C. S. (1999), Achievable benchmarks of care: the ABCTMs of benchmarking. *Journal of Evaluation in Clinical Practice*, 5: 269–281. doi:10.1046/j.1365-2753.1999.00203.

4 STAR RATING METHODOLOGY DEVELOPMENT

The CY 2016 PFS rule finalized the ABCTM benchmark as the anchor for 5-star rating, but stipulated that the technique used to categorize the remaining reporters into 1- to 4-star ratings would be established on a sub-regulatory basis through analysis of the data and feedback from stakeholders and experts. The Physician Compare support team worked closely with the TEP to develop a star rating methodology based on the ABCTM benchmark. Section 4.1 describes the details of how the ABCTM benchmark is calculated. Section 4.2 outlines the desired features of a successful star rating strategy and identifies two candidate methods that have those features. Section 4.3 presents the analyses developed to assess how reliably each method assigns reporters to star ratings. Section 4.4 discusses the removal of outliers from performance distributions prior to calculating star rating cut-offs. Section 4.5 compares two candidate star rating methodologies. Section 4.6 communicates CMS's decision on the star rating methodology.

4.1 ABCTM Benchmark Calculation

The ABCTM benchmark will be used as the threshold for outstanding performance for each quality measure/submission mechanism combination. The ABCTM benchmark represents the pared mean performance score among the highest performing reporters that treated at least 10 percent of the eligible patient population. In order to calculate the benchmark, one must first determine the adjusted performance fraction (APF). The APF is calculated by multiplying a reporter's performance score by a beta-binomial adjustment factor to help ensure the benchmark is not overly influenced by high performance rates that occur by chance due to low sample sizes. The beta-binomial adjustment was selected over the Bayesian adjustment, as it is preferable to adjust performance rates toward the average performance rate for a given measure as opposed to a performance rate of 50 percent. Reporters are rank-ordered by their APF and the top performing reporters that treated at least 10 percent of the total eligible patient population are selected. Then, the pared mean of the raw performance scores for this top subset of reporters is calculated. The result is the ABCTM benchmark, which serves as the 5-star cut-off for each measure. Any reporter with a performance score equal to or higher than the ABCTM benchmark will receive a rating of five stars for that particular measure. Benchmarks are subsequently tested to ensure they meet all reliability criteria. This process is detailed later in Section 4.3.

4.2 Star Rating Assignment with Candidate Methods

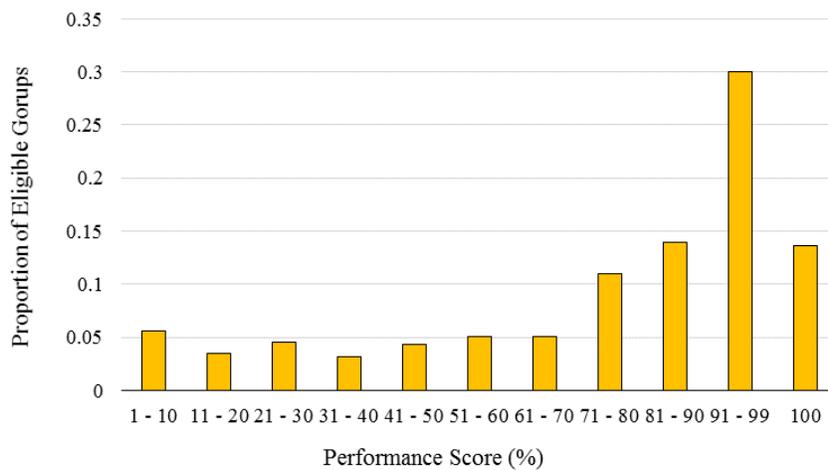
To devise a method to assign reporters into 1- to 4-star ratings, the Physician Compare support team assembled a set of characteristics desirable in a star rating methodology, based on mandated statistical standards and input from TEP members, other stakeholders, and CMS. The optimal technique should:

- Produce 1- to 4-star ratings that are relative to the threshold set by the ABCTM benchmark;
- Avoid a forced distribution by accommodating each measure/mechanism combination’s unique observed performance distribution;
- Not restrict the number of reporters who can receive a 5-star rating;
- Not unduly impede achievement of a moderate to good star rating; and,
- Generate meaningful and statistically reliable benchmarks.

Multiple methodologies were tested and iteratively adapted in an effort to specify a star attribution strategy that exhibits these predefined attributes. Two techniques remained strong candidates for the Physician Compare Star Rating Methodology: the Equal Ranges and Cluster methods. The Equal Ranges method establishes the 1- to 4-star cut-offs by dividing the performance scale between the minimum reported performance score and the ABCTM benchmark into four equal ranges. Alternatively, to assign reporters to the 1- to 4-star categories, the Cluster method first removes 5-star performers who exceeded the ABCTM benchmark for the given measure and then uses a clustering algorithm to establish cut-offs that maximize the similarity of the performance scores within each star rating group and minimize between-group similarity.

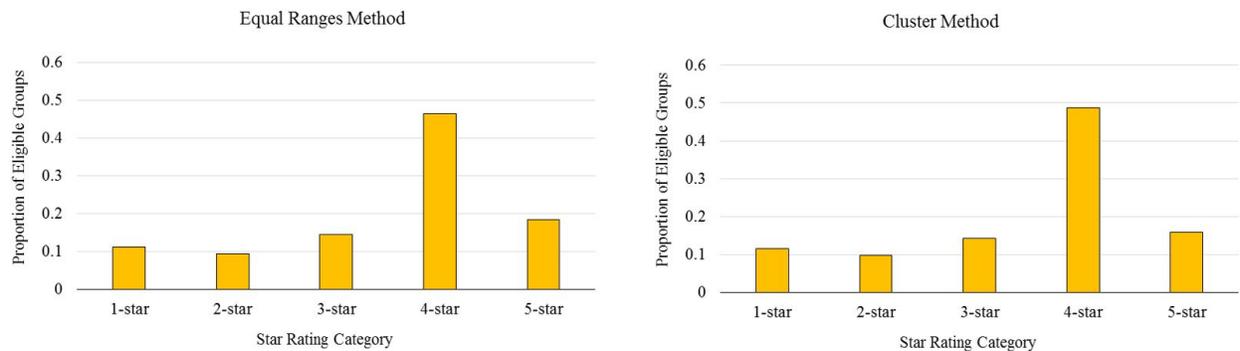
During the TEP meeting in February 2017, the Physician Compare support team presented the star rating allocations for sample measures with a variety of performance score distributions. Figure 1 displays the left-skewed performance score distributions for the group registry data submitted for PQRS 47 – Advanced Care Plan.

Figure 1. PQRS 47 – Advanced Care Plan Group Registry Performance Rate Distribution



The resulting star rating distributions generated by both candidate methods mirrored the left-skewed distribution of the data for this measure/mechanism combination where most groups had a performance score of higher than 70 percent. In this example, over 70 percent of groups achieved three or more stars, regardless of the candidate star rating method used (Figure 2). A similar reflection of the underlying performance rate distribution was consistently observed across a multitude of distribution types (e.g. bimodal, u-shaped, skewed, normal).

Figure 2. Star Rating Distributions – PQRS 47 – Advanced Care Plan Group Registry Data



After presented with star rating distributions of alternative star rating methods, the TEP recommended that the Physician Compare support team continue investigating the Equal Ranges and Cluster methods because neither method forced a distribution, nor was either overly punitive for a range of measures with varying distributional properties, as exemplified in Figure 2. After establishing that these two methods meet the predefined criteria for star rating attribution, the Physician Compare support team conducted extensive testing to ensure that the methods could reliably assign reporters to star rating categories. Additionally, stakeholder outreach was conducted to obtain additional feedback on the two options.

4.3 Star Rating Reliability Analyses

All data publicly reported on Physician Compare must be reliable. To meet this public reporting requirement for star rating categories, the Physician Compare support team must confidently assess that (1) the star-category-defining cut-offs are robustly calculated and (2) reporting entities are being reliably assigned a star rating, given the precision of the available data. Robust star rating cut-offs are precise, meaning that the cutoffs remain stable across expected data fluctuations. Measure/mechanism combinations with a low number of reporters are less likely to produce reliable cut-offs than those with higher numbers of reporters. Since reliable cut-offs tend to be calculated from measure/mechanism combinations with more reporters, they are likely to be stable over time. Additionally, for each measure/mechanism combination, the precision of the reported performance scores supports reliable classification of reporters into star categories. A reporter’s performance score will be more precise when the entity submitted data

for a larger number of patients. If the performance scores used to assign reporters to star categories for a given measure and mechanism are based on low patient sample sizes, reporters are more likely to be assigned to a particular star category based on chance. Because these unreliable star ratings would not provide meaningful comparisons between the reporters they score, they should not be used for public reporting.

Multiple characteristics of the performance data (most notably, the number of reporters and the variation in performance rates across reporters) must align to support star ratings that meaningfully differentiate between individual reporters. However, inability to create meaningful high-level categories from measure data does not inherently speak to the reliability of the measure data itself. If the performance scores pass the measure data reliability tests described in Section 5.2.2, the individual performance scores are considered meaningfully different. Therefore, measures data that do not pass these stringent star assignment reliability tests could still potentially be appropriate for public reporting in the downloadable database, or on public-facing profile pages as percentages instead of as star ratings, as long as the data meet all other public reporting criteria.

As recommended by the TEP, the star rating reliability standards described throughout Section 4.3 were chosen conservatively. The next two sections describe the metrics of reliable star assignment used by the Physician Compare support team, as well as the details behind the cut-off robustness (Section 4.3.1) and performance rate precision (Section 4.3.2) tests.

4.3.1 Cut-off Robustness

To assess the reliability of the star rating cut-off themselves, the Physician Compare support team tested how stably each method calculates 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 amount of variation in performance across reporting entities. To investigate each method's capacity to produce reliable cut-offs, a bootstrapping analysis was conducted where reporters' performance scores were randomly sampled with replacement (e.g. adding the randomly sampled reporter back to the sampling pool after each iteration) until the sample size was equal to the number of reporters for that measure. This process was repeated 500 times for each measure/mechanism combination. For each simulated data set, the star rating cut-offs were recalculated and the simulated cut-offs were used to reassign each reporter to a simulated star rating category, using their original performance score.

The TEP concurred that cut-off robustness is an important concern when determining star rating reliability and agreed that this analysis would be helpful to determine which measures

have reliable cut-offs. Thus, to assess the reliability of the simulated star rating cut-offs, the Physician Compare support team counted the number of times the simulated star rating matched the assigned star rating based on the observed performance distribution. For calculated cut-offs to be considered reliable, the simulated and observed star ratings must demonstrate 80 percent accuracy (e.g. they must match 80 percent of the time). Further, it is more undesirable for a star rating methodology to misclassify a reporter by two or more stars than by just one star. Certain star rating classifications could have high accuracy but still have a concerning percentage of multi-star movement. To limit the number of potential multi-star misclassifications, star attribution was deemed reliable if its simulated star ratings are only off by two or more stars less than five percent of the time. Additionally, the Fleiss' kappa was calculated to compare the inconsistency across simulated data sets to what would be expected to exist due to chance. A star rating scheme must have a kappa of at least 0.6 to be considered reliable. Finally, the absolute value of movement along the performance rate scale for each cut-off (including the ABCTM benchmark) was assessed to ensure that no simulated star rating cut-offs fluctuated more than five percentage points away from the observed star rating cut-offs. Measure/mechanism combinations with a lower number of reporters are more likely to violate these cut-off reliability thresholds. Importantly, if the data for a measure/mechanism combination fails to meet any one of these criteria, the measure/mechanism combination would not be considered for publication on public-facing profile pages as a star rating display. The TEP agreed with the thresholds presented to them and recommended moving forward with these star rating reliability standards.

Contingency tables were used to display the accuracy of the simulated star rating assignments compared to the actual star rating assignments, as shown using PQRS 47 – Advanced Care Plan group registry data in Figure 3. The percentages highlighted in blue display the accuracy for each star rating category. For example, the 42 groups who were originally categorized as 1-star groups were classified into the simulated 1-star category 100 percent of the time. For this measure/mechanism combination, the simulated star ratings matched with their respective observed star ratings more than 99 percent of the time. Fleiss' kappa for the PQRS 47 – Advanced Care Plan group registry data star rating cut-off simulations was 0.99. Because the star assignments for this measure/mechanism combination exhibit high accuracy (≥ 80 percent), little multi-star movement (≤ 5 percent), and an adequate Fleiss' kappa (≥ 0.6), PQRS 47 – Advanced Care Plan group registry data would remain eligible for public reporting on the Physician Compare website, given that all other criteria for eligibility are also met.

Figure 3. Star Rating Cut-off Reliability Contingency Table – PQRS 47 – Advanced Care Plan Group Registry Data

			Counterfactual Stars				
			1-star	2-stars	3-stars	4-stars	5-stars
Actual Stars	1-star	n = 42	100%	0%	0%	0%	0%
	2-stars	n = 35	0%	100%	0%	0%	0%
	3-stars	n = 54	0%	0%	100%	0%	0%
	4-stars	n = 173	0%	0%	0%	99%	1%
	5-stars	n = 69	0%	0%	0%	0%	100%

Conversely, the star rating scheme for PQRS 370 – Depression Remission at Twelve Months group Web Interface data (Figure 4), had 74.1 percent accuracy, 1.7 percent multi-star movement, and a kappa of only 0.56.

Figure 4. Star Rating Cut-off Reliability Contingency Table – PQRS 370 – Depression Remission at Twelve Months Group Web Interface Data

			Counterfactual Stars				
			1-star	2-stars	3-stars	4-stars	5-stars
Actual Stars	1-star	n = 15	85%	15%	0%	0%	0%
	2-stars	n = 8	8%	79%	12%	1%	0%
	3-stars	n = 8	0%	16%	55%	22%	6%
	4-stars	n = 2	0%	0%	4%	35%	61%
	5-stars	n = 1	0%	0%	0%	0%	100%

Compared to the PQRS 47 – Advanced Care Plan results highlighted in Figure 3, the contingency table in Figure 4 reflects a much lower frequency of accurate reclassification of star ratings in the bootstrapping analysis. This implies that the star rating cut-offs for PQRS 370 – Depression Remission at Twelve Months group Web Interface data were less robust across simulations, which is likely related to the low number of reporters for the measure. Because robust cut-offs could not be established for the PQRS 370 group Web Interface data, the star ratings would be considered unreliable and unfit for public reporting as a star display.

4.3.2 Performance Rate Precision

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 were not influenced heavily by random error, each reporter’s actual

performance rate and patient population sizes were used to simulate counterfactual performance rates using a binomial distribution. The ABCTM benchmark and the star rating cut-offs were recalculated for each simulation, and reporters were 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 was determined. Similar to the cut-off reliability assessment, a reliable star rating scheme will have at least 80 percent accuracy, 5 percent or less multi-star movement, and a Fleiss' kappa of 0.6 or more. The TEP similarly agreed with applying these thresholds for assessing performance rate precision and recommended moving forward with these star rating reliability standards.

PQRS 110 – Influenza Immunization group Web Interface is an example of a measure/mechanism combination that passes all standards for performance rate precision reliability (Figure 5). Across simulations, the star rating scheme for these data demonstrated 88.1 percent accuracy, less than 0.1 percent multi-star movement, and a kappa of 0.74. Therefore, the star rating cut-offs for PQRS 110 – Influenza Immunization group Web Interface data would be considered reliable.

Figure 5. Star Rating Assignment Reliability Contingency Table – PQRS 110 – Influenza Immunization Group Web Interface Data

			Counterfactual Stars				
			1-star	2-stars	3-stars	4-stars	5-stars
Actual Stars	1-star	n = 13	83%	17%	0%	0%	0%
	2-stars	n = 50	3%	86%	10%	0%	0%
	3-stars	n = 120	0%	4%	90%	5%	0%
	4-stars	n = 70	0%	0%	14%	85%	1%
	5-stars	n = 16	0%	0%	0%	7%	93%

PQRS 334 – Appropriate Use of CT Scan for Chronic Sinusitis group registry data were unable to support reliable star assignments, given the precision of the constituent performance rates (Figure 6). The star rating scheme for PQRS 334 – Appropriate Use of CT Scan for Chronic Sinusitis group registry data showed only 75.4 percent accuracy, 3.1 percent multi-star movement, and a kappa of 0.50. The star assignments for PQRS 334 – Appropriate Use of CT Scan for Chronic Sinusitis fail multiple reliability requirements. Thus, a star rating display would not be recommended for this measure/mechanism combination on Physician Compare.

Figure 6. Performance Rate Precision Reliability Contingency Table – PQRS 334 – Appropriate Use of CT Scan for Chronic Sinusitis Group Registry Data

			Counterfactual Stars				
			1-star	2-stars	3-stars	4-stars	5-stars
Actual Stars	1-star	n = 3	100%	0%	0%	0%	0%
	2-stars	n = 2	37%	29%	23%	9%	2%
	3-stars	n = 14	4%	12%	47%	29%	8%
	4-stars	n = 41	0%	1%	13%	72%	14%
	5-stars	n = 40	0%	0%	2%	9%	89%

4.4 Removal of Major Outliers

The star rating cut-offs for a given measure and mechanism should not be heavily influenced by a small number of reporting entities with outlying performance rates, which was a consideration the TEP asked the Physician Compare team to investigate further. To address these concerns, the Physician Compare support team studied the effect of removing performance outliers on the star ratings produced by the candidate methods (e.g. Equal Ranges and Cluster). The star rating cut-offs change substantially after removing major outliers³ for measures with largely left-skewed distributions (e.g. for high performing measures), regardless of the candidate method used for determining the cut-offs. Non-high-performing measures were unaffected because those data did not contain major outliers that impacted the cut-off values. Figure 7 displays the performance score distribution for PQRS 324 – Appropriate Use of Cardiac Stress Imaging, an example of a high performing measure. Before removing major outliers for PQRS 324 – Appropriate Use of Cardiac Stress Imaging group registry data, the 2-, 3-, and 4-star cut-offs as determined by the Equal Ranges method were 75 percent, 83 percent, and 92 percent, respectively (Figure 8). After major outliers were removed, the 2-, 3-, and 4-star cut-offs became 96 percent, 98 percent, and 99 percent, respectively (Figure 9).

While these cut-offs are close together, they are more sensible because they are based on the portion of the performance score distribution with the majority of the data. Removing outliers before assigning star ratings ensures the cut-offs used are not influenced by non-representative, stand-alone performance scores that could be unstable over time. For high performing measures (e.g. measures for which the vast majority of clinicians and groups report a high performance score), after removing outliers, the star cut-offs themselves remained reliable, whereas the ability

³ Major outliers are identified using the Tukey's fences method, which is based on the interquartile range (IQR). If Q1 (25th percentile) is the lower quartile of the performance rate distribution, a major outlier is defined as any observed performance rate lower than three times the IQR below Q1.

to categorize reporters meaningfully into the star rating categories based on performance rate precision was greatly diminished because the cut-offs were close together (so even reasonably precisely measured rates might span multiple cut-offs). Thus, measures that exhibit this behavior do not meet the Physician Compare public reporting standards required for reporting measures with a star rating display on public-facing profile pages.

Figure 7. PQRS 324 – Appropriate Use of Cardiac Stress Imaging Group Registry Raw Performance Score Distributions

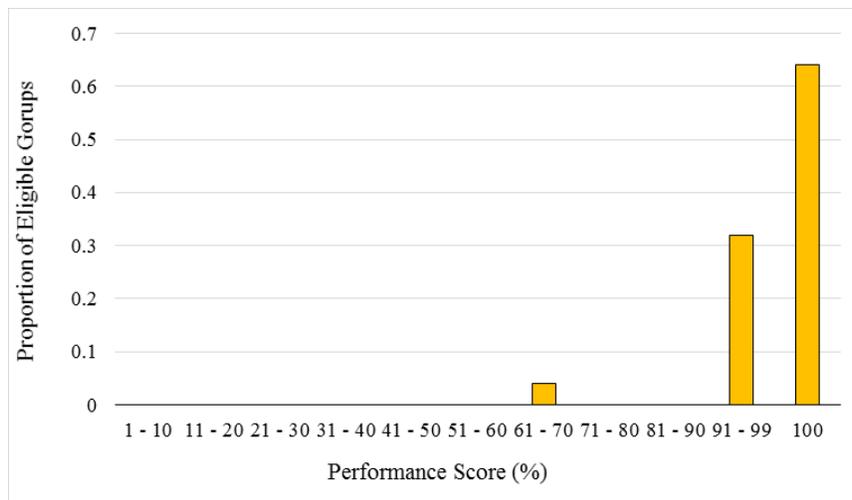
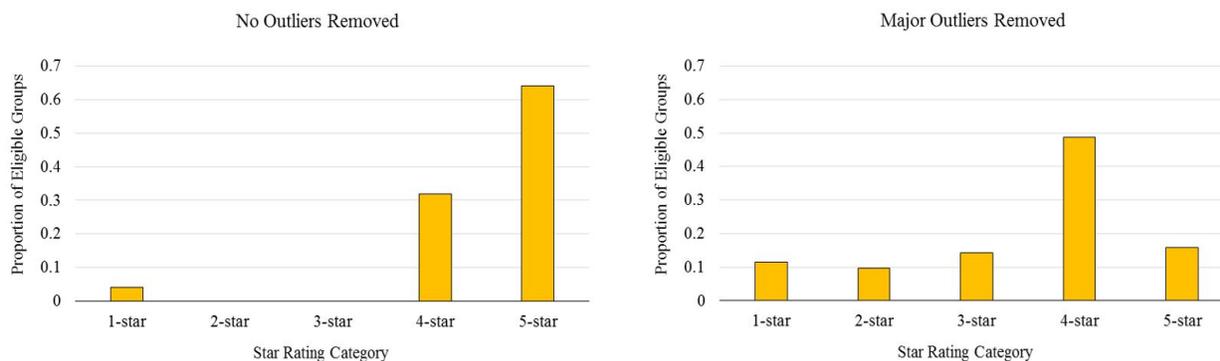


Figure 8. Star Rating Distributions – PQRS 324 – Appropriate Use of Cardiac Stress Imaging with and without major outliers removed



TEP Input on Outliers

The TEP members recommended removing outliers and reinforced that the cut-offs should be created using the more data-dense areas of the performance score distributions. They also supported excluding high performing measures for which reporters cannot be meaningfully categorized. One member suggested that the Physician Compare support team recognize reporters receiving a 5-star rating on measures for which the benchmark and 5-star reporters can

be reliably determined but the remaining reporters cannot be classified into the 1- through 4-star rating categories. Reporters exceeding the benchmark would receive an indication of high performance on their public-facing profile pages, whereas the remaining reporters would have no indication of performance on their profiles. During a series of webinars, other stakeholders expressed strong support for this idea because it would acknowledge reporters who excelled at particular aspects of health care.

4.5 Comparison of Candidate Star Rating Methods

To help guide the selection of a final methodology, star ratings were produced for many measure/mechanism combinations using PY 2015 data after removing outliers. Both star rating reliability analyses (cut-off robustness and performance rate precision) were performed on the star rating assignments. The Equal Ranges method was able to generate reliable star ratings for 23 percent more measure/mechanism combinations than the Cluster method. Based on this finding, it appears that the Equal Ranges method can produce reliable star ratings for a more diverse set of performance score distributions. Because the Cluster method relies on maximizing the distance between four clusters within an available performance distribution, it depends on the stability of overall distribution across simulations to produce consistent star categorizations. Conversely, the Equal Ranges method only requires that the distance between the ABCTM benchmark and the minimum non-outlier data point remain similar across simulations to generate consistent star categorizations. These results provide evidence that the Cluster method is more sensitive to minor changes in the entire performance distribution than Equal Ranges.

4.5.1 TEP Input on the Decision between Candidate Methods

The TEP members recommended the Equal Ranges method and remarked that it seems to categorize reporters more reliably. One TEP member commented that, because it generates more stable cut-offs, using the Equal Ranges method would give reporters a better sense for what performance score to strive for over time, compared to the Cluster method. Because the specifications for the current PQRS process measures align with the Quality Payment Program quality measures available for the PY 2017 performance period, this characteristic is relevant even as public reporting moves from PQRS to the Merit-based Incentive Payment System (MIPS). It is also invaluable to reporters who continue to seek improvement in future performance periods under MIPS. Furthermore, the method generates reliable star ratings for a larger number of measure/mechanism combinations, which would ultimately result in more data being available to website users over time.

4.6 Star Rating Methodology Decision

Based on the analyses performed by the Physician Compare support team and the feedback from TEP members and stakeholders, CMS decided to move forward with the Equal Ranges method for star 1-to 4-star rating attribution. The technique produces reliable, stable star ratings with robustly calculated cut-offs. The resulting star rating distributions naturally reflect the underlying performance rate distribution, which enhances website user comprehension of the differences in star rating categories and avoids forced assignments. The 5-star ABCTM benchmark establishes top performers based on the currently available data, provides a point of comparison, and represents quality while being realistically achievable in the current clinical climate. The Equal Ranges method outperforms the Cluster method in its ability to produce reliable cut-offs and assign reporters to star categories, given the precision of the available performance scores. Using the ABCTM benchmark as the anchor and the Equal Ranges method to attribute 1- to 4-stars, CMS will restart its phased approach to public reporting by a subset of group-level PY 2016 PQRS measures available for public reporting as star ratings in late 2017.

5 PY 2016 MEASURE SELECTION PROCESS

Physician Compare quality measures are made public via a downloadable database and public-facing profile pages. All measure/mechanism combinations that are accurate, valid, and reliable (as per the public reporting standards outlined in Section 3) are eligible for publication in the downloadable database. CMS will display star ratings on profile pages for a subset of group PY 2016 PQRS measures in late 2017. Additionally, performance scores for non-PQRS QCDR measures (available for both groups and individuals in PY 2016) and CAHPS for PQRS measures, which were not eligible for star rating display per the CY 2016 PFS final rule, will appear on public-facing profile pages as percentages. Any measures posted to profile pages must also be understood by and resonate with website users as determined through user testing.

The remainder of Section 5 describes the PY 2016 measures selection process and recommendations for measure publication presented to the TEP on August 28, 2017. Section 5.1 outlines the criteria that establish which measure data are available for publication on Physician Compare. Section 5.2 details the analyses undertaken to determine which of the available measure data meet Physician Compare's public reporting standards. Section 5.3 presents CMS' plan for Physician Compare PY 2016 quality measure publication, which unfolds in late 2017.

5.1 Eligible Measures Data

CMS conservatively decided against publicly publishing certain PY 2016 quality measures *a priori*. Section 5.1.1 details these measure-level exclusions from Physician Compare public reporting. Further, for each measure/mechanism combination, the Physician Compare support team requires that data submitted by individual reporters meet certain requirements to be included in our analyses. Section 5.1.2 reports the eligibility criteria for individual reporters.

5.1.1 Excluded Measures

Certain measures are ineligible for public reporting based on CMS guidance, and were therefore excluded from analyses testing compliance with Physician Compare public reporting standards (see Section 1). Specifically, these measures are those offered exclusively through a part of a measures group, all continuous measures (e.g. "Median Time from Emergency Department (ED) Arrival to ED Departure for All Discharged ED Patients"), all non-risk adjusted outcome measures (see Section 5.2.3), and all new measures available for public reporting for the first time in PY 2016. Furthermore, certain measures were affected by updates to the International Classification of Disease, 10th Revision (ICD-10); these measures were also excluded *a priori*. Additionally, to ensure that a measure/mechanism's data were adequately powered to support statistically sound conclusions, all measure/mechanism combinations with fewer than 20 reporters were ineligible for public reporting. In total, 350 group PY 2016 PQRS

measure/mechanism combinations and 461 individual PY 2016 PQRS measure/mechanism combinations were ineligible for further analyses and publication on Physician Compare in late 2017 due to one of these exclusionary criteria. After the identification of ineligible measure/mechanism combinations, the Physician Compare support team identified 260 group-level measure/mechanism combinations and 504 individual-level measure/mechanism combinations eligible for further testing.

5.1.2 Excluded Reporters

Within each measure/mechanism combination, individual reporters must meet certain requirements to remain eligible for compliance testing. For a reporter's measure data to be eligible for public reporting, the reporter must meet the minimum reporting requirements for the relevant submission mechanism. The 2016 EHR quality measure data are not being publicly reported on Physician Compare this year. CMS will not be adding this additional submission mechanism in the final year of the PQRS program as these data have not been previously published on the website. Beginning with the first year of Quality Payment Program data, all submission mechanisms will be available for public reporting, if technically feasible and if the data meet public reporting criteria (Section 1). Therefore, EHR entries were excluded from consideration for PY 2016 public reporting. Further, reporters must have submitted performance data for at least 20 eligible beneficiaries and earned higher than a zero percent performance score (or lower than 100 percent on an inverse measure) for the measure data to be considered for analysis. As a quality assurance check, a reporter must not have missing or nonsensical performance or reporting data (e.g. a numerator exceeding a denominator). If data submitted by a reporter do not meet these criteria, the reporter is eliminated from public reporting consideration for the relevant measure.

5.2 Meeting Public Reporting Standards

CMS has contracted the Physician Compare support team to ensure the compliance of any publicly reported measure data with mandated public reporting standards, which require that published data are accurate, reliable, valid, and comparable across available submission mechanisms. All measures published to public-facing profile pages must also be understood by and resonate with consumers. The remainder of this section outlines the Physician Compare support team's testing for PY 2016 measure data eligible for publication on Physician Compare in late 2017.

5.2.1 Comparability

Comparing the performance score distributions from different submission mechanisms for the same measure shows that raw performance data submitted through distinct submission mechanisms are not always comparable. Thus, as per TEP recommendation from PY 2015, all of

the following analyses are performed for each measure/mechanism combination, rather than aggregating data from different mechanisms. When deciding which submission mechanism data to report publicly, the Physician Compare support team determines which compliant submission mechanism⁴ had a higher number of reporters. For group submissions, the total number of individual clinicians contributing to the groups that reported through the relevant mechanism is counted to make this determination.

5.2.2 Reliability

To assess reliability of candidate measure data, the Physician Compare support team performs an exhaustive series of analyses, including tests to establish performance rate reliability and, for measures eligible for star rating displays on profile pages, confirm reliable star rating assignments. The remainder of this section outlines the approach for establishing measure-level and star rating-level reliability for PY 2016 measures.

Measure Data

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. Statistically, reliability depends on performance variation for a measure across reporters, the random variance in performance for a measure within a provider's panel of attributed beneficiaries, and the number of beneficiaries attributed to the provider. High reliability for a measure suggests that comparisons of relative performance across reporters are likely to be stable over different performance periods and that the performance of one reporter on the quality measure can be confidently distinguished from another. Potential reliability values range from zero to one, where one (highest possible reliability) means that all variation in the measure's rates is the result of differences in performance across reporters, while zero (the lowest possible reliability) means that all variation is a result of measurement error.⁵ Based on prior TEP recommendation, the Physician Compare support team assesses the reliability of the performance data submitted through all available submission mechanisms for each measure. Reliability is calculated using two methods: (1) the beta binomial test and (2) a split half reliability test. If a measure/mechanism combination passes these reliability tests, the performance scores are considered meaningfully different across reporting entities.

For the first test, a beta binomial model is fit to each measure's performance score data, then parameters from the simulated beta binomial models are used to calculate reporter-to-reporter variance and error for each measure. Using the reporter-to-reporter variance, binomial

⁴ Compliant submission mechanisms are ones that have passed reliability and validity testing

⁵ For more information about reliability testing for clinician performance measurement, see "Reliability of Provider Profiling: A Tutorial" by John Adams, RAND. http://www.rand.org/pubs/technical_reports/TR653.html

error, and observed performance scores for each measure, a reliability score for each reporter is calculated. The Physician Compare support team uses the 25th percentile of the reliability scores for a given measure to determine whether its data are reliable. Literature suggests that groups and individuals should have different thresholds for adequate reliability (0.7-0.8 for groups and 0.9 for individuals).⁶ Because groups using a registry could technically be comprised of only two clinicians, the Physician Compare support team uses the individual standard (0.9) for these reporters. Web Interface users must be groups with at least 25 clinicians, so the upper limit of the accepted range for sufficient group reliability (0.8) was adopted as the cutoff. Thus, if the 25th percentile of the reliability scores determined by the beta binomial model for a given measure was greater than or equal to 0.9 for individual clinician submissions or groups submitting data through a registry, or was greater than or equal to 0.8 for groups submitting data through the Web Interface, the measure data was considered reliable.

The second reliability assessment, the split half reliability test, randomly divides the population into halves and compares performance between the two halves. If the resultant intraclass correlation coefficient, which quantifies the resemblance of the two randomly selected halves of the dataset, was greater than or equal to 0.75, the measure was considered reliable. A measure must pass both the beta binomial test and the split half reliability test to maintain candidacy. The Physician Compare support team identified 78 group-level measure/mechanism combinations and 150 individual-level measure/mechanisms with reliable measure data.

Star Rating Assignment

As discussed in Section 4.3, to be considered for public reporting on public-facing group profile pages as a star display, the Physician Compare support team must confidently ensure that (1) the star-category-defining cut-offs are robustly calculated and (2) reporting entities are being reliably assigned a star rating, given the precision of the available data. Star rating assignments for the group measures eligible for star rating displays were determined. Measure/mechanism combinations with star ratings passing the star rating reliability testes described in Section 4.3 were eligible for public reporting, given compliance with other public reporting standards.

5.2.3 Validity

Validity reflects the degree to which a metric measures what it purports to measure.⁷ The Physician Compare support team assesses each quality measure's validity, which, in this context, refers to the extent to which the observed performance rates on measures were impacted by factors unrelated to true performance, such as characteristics of the reporter's patient population

⁶ Hays RD, Revicki D. Reliability and validity (including responsiveness). In: Fayers P, Hays R, eds. *Assessing Quality of Life In Clinical Trials*. New York: Oxford University Press; 2005

⁷ Moskal, B.M., & Leydens, J.A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). <http://pareonline.net/getvn.asp?v=7&n=10>

(e.g. case-mix) or selective reporting on a non-representative sub-segment of the patient population. The validity of measure data is addressed in two ways: (1) identification of non-risk adjusted outcome measures and (2) an investigation into the possibility of selective reporting.

Non-Risk Adjusted Outcome Measures

Performance on certain measures can be influenced by the patient population served by a clinician. Unlike process measures, which portray the frequency with which a clinical standard of care was performed by a clinician, outcome measures reflect patient health (e.g. the percentage of diabetic patients with controlled blood sugar). These measures are supposed to show whether the care provided by clinicians translates into positive health outcomes. However, since PQRS outcome measures do not include risk adjustment strategies, clinician performance on these measures could be associated with the types of patients treated, given that certain patient populations will be healthier than others (e.g. a clinician serving patients with more complex clinical conditions might have worse health metrics than one treating less complicated patients). If characteristics of the patient population impact the observed performance rates on outcome measures, those rates would be an invalid representation of the quality of care provided by the clinician. In previous program years, when the Physician Compare support team controlled for the demographic characteristics and clinical conditions of patients included in outcome measure data, case-mix effects on performance were observed. Therefore, CMS, with strong TEP support, has chosen to exclude all PY 2016 non-risk adjusted outcome measures from public reporting eligibility in late 2017.

Selective Reporting

Reporters using claims, qualified registries, or QCDRs are only required to submit quality data on half of their eligible patients to be considered satisfactory reporters. As the TEP pointed out in PY 2015, this creates an opportunity for savvy reporters to select their best patients on whom to submit data in an attempt to inflate their performance scores. If a subset of reporters cherry-picked patients for measure data submissions, their artificially elevated performance rates would be an invalid comparator to those reporters choosing a more inclusive representation of their patient population.

To address this concern, the Physician Compare support team runs a Pearson correlation between the reporting rates (e.g. the portion of eligible patients reported) and the performance rates for each measure. For a measure for which some reporters selectively reported their better-performing patients, reporters with lower reporting rates would be expected to have better performance than their counterparts who submitted a more representative sample (e.g. reporting and performance rates would be inversely related). Thus, if the correlation is statistically significant and the correlation between the reporting rates and performance rates is less than or

equal to -0.6, the measure/mechanism combination is eliminated due to evidence consistent with selective reporting.

5.3 Selected Measures

After assessing adherence of PY 2016 quality measures with Physician Compare public reporting standards, the Physician Compare support team presented the compliant measure/mechanism combinations to CMS for review. CMS considered analytic results and feedback from stakeholders and the TEP when deciding on the final set of measures to publish. All measures with compliant data will be included in the Physician Compare downloadable database, whereas a subset of compliant measures will be published on public-facing profile pages. The remainder of this section outlines the considerations for measure publication in the downloadable database and on public-facing profile pages.

Downloadable Database

The downloadable database serves as an exhaustive compilation of all measures meeting the mandated statistical standards, including non-PQRS measures published to QCDR websites.⁸ It details the performance scores received by reporters submitting quality data for each measure, the number of patients on whom they reported, and an indication of the submission mechanism used. For a measure to be included in the downloadable database, it must have at least one submission mechanism whose data pass the eligibility criteria established in Section 4.1 and the compliance testing detailed in Section 4.2. If a particular measure has multiple submission mechanisms with compliant data, the Physician Compare support team selects the most represented mechanism using the rules outlined in Section 5.2.

Public-Facing Profile Pages

In addition to the statistical public reporting standards, measures published to public-facing profile pages must be understood by and resonate with website users. Past testing of plain language versions of measure titles and specifications indicate that users are able to comprehend and make actionable decisions from the available measures. Performance scores for non-PQRS QCDR and CAHPS for PQRS measures will be published on public-facing profile pages as percentages if they meet all public reporting standards required of live-site measures. Regarding PQRS measures, CMS will restart its phased approach to public reporting by publicly reporting just a small sub-set of group-level measures reported as star ratings (using the Equal Ranges method anchored by the ABCTM benchmark described in Section 3). Appendix A details the measures to be published on public-facing profile pages, including information about the star

⁸ QCDRs can choose to publish their non-PQRS measures on their own websites. If a QCDR opts to publish non-PQRS measures on its own website, those performance data will still be included in Physician Compare's downloadable database but will not appear on public-facing profile pages. Instead, a link to the QCDR's website will appear on the relevant profile pages.

rating cut-offs and percentage of groups in each star rating category. Appendix B displays the performance rate distributions for these same measures.

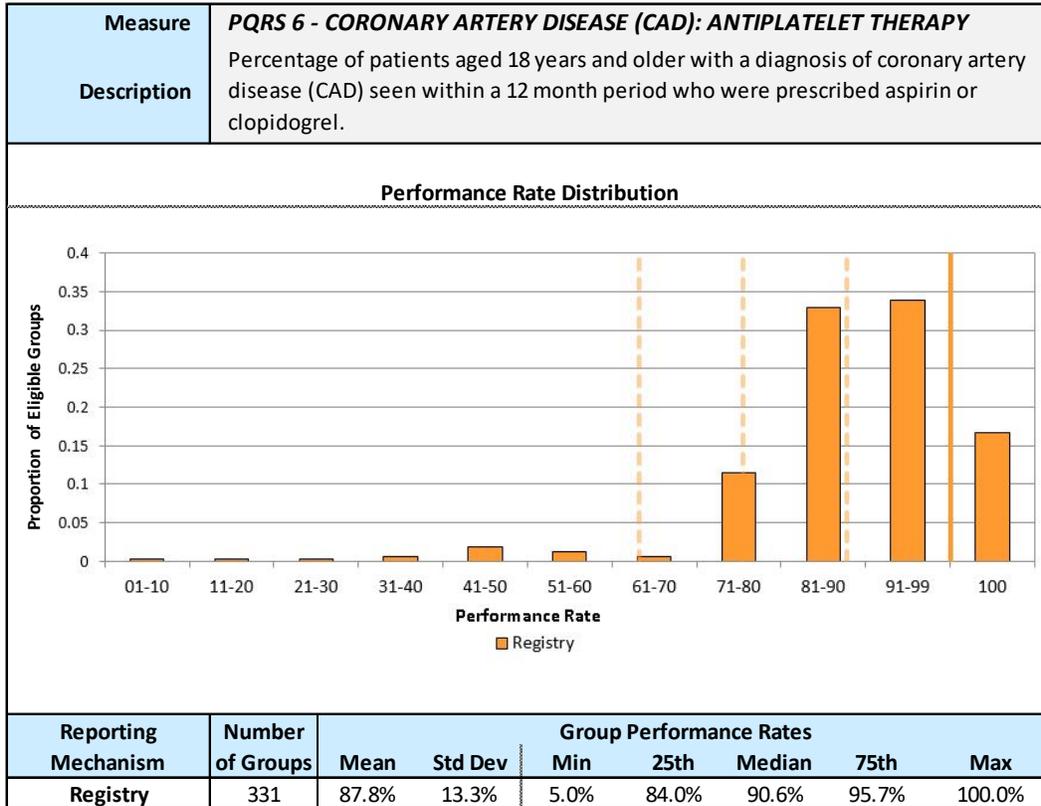
APPENDIX A – STAR RATING CUT-OFFS AND REPRESENTATION

The following table presents the total number of reporters, number of reporters by star value, and star cut-off values for each measure that met all PQRS public-facing profile display public reporting requirements for publication in late 2017.

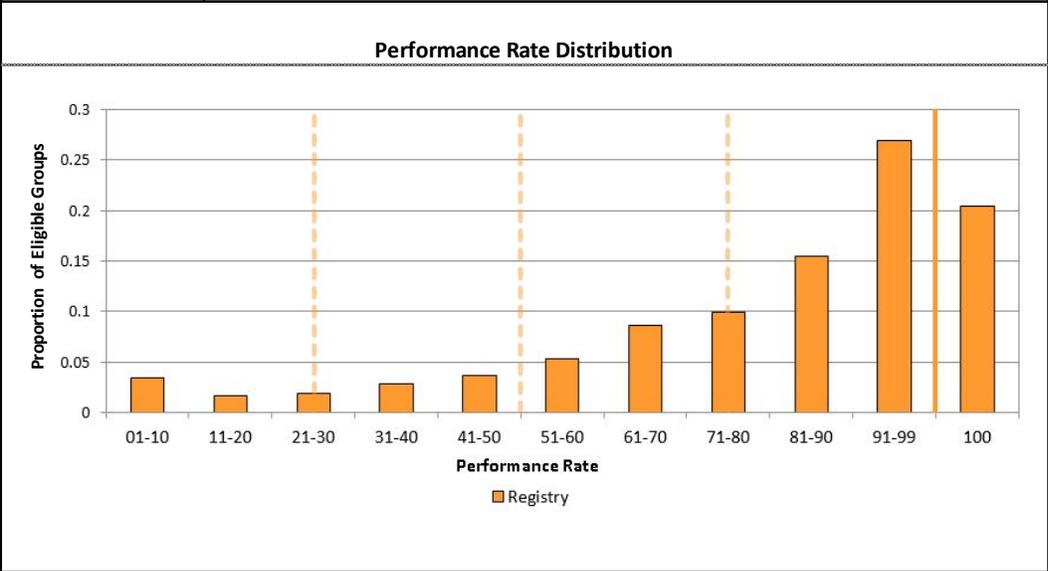
#	Measure Information Title	Reporting Mechanism	# of Groups	Star Rating Cut-offs			
				2	3	4	5
6	Coronary Artery Disease (CAD): Antiplatelet Therapy	Group Registry	331	61%	74%	87%	100%
47	Advance Care Plan	Group Registry	677	25%	50%	75%	100%
48	Urinary Incontinence: Assessment of Presence or Absence of Urinary Incontinence in Women Aged 65 Years and Older	Group Registry	128	25%	50%	75%	100%
51	Chronic Obstructive Pulmonary Disease (COPD): Spirometry Evaluation	Group Registry	44	24%	48%	72%	96%
110	Preventive Care and Screening: Influenza Immunization	Web Interface	264	24%	45%	66%	87%
111	Pneumonia Vaccination Status for Older Adults	Web Interface	264	25%	48%	71%	93%
113	Colorectal Cancer Screening	Web Interface	255	21%	42%	63%	84%
117	Diabetes: Eye Exam - National Quality Strategy	Web Interface	237	21%	40%	60%	80%
128	Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan	Web Interface	270	35%	54%	74%	93%
134	Preventive Care and Screening: Screening for Clinical Depression and Follow-Up Plan	Web Interface	253	23%	47%	70%	93%
226	Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention	Group Registry	1,659	63%	75%	87%	100%
238	Use of High-Risk Medications in the Elderly	Group Registry	654	60%	74%	87%	100%
318	Falls: Screening for Future Fall Risk	Web Interface	261	25%	50%	74%	99%

APPENDIX B – GROUP-LEVEL PQRS PERFORMANCE RATE HISTOGRAMS

The following histograms present the performance rate distributions for each measure that met all PQRS public-facing profile display public reporting requirements for publication in late 2017.

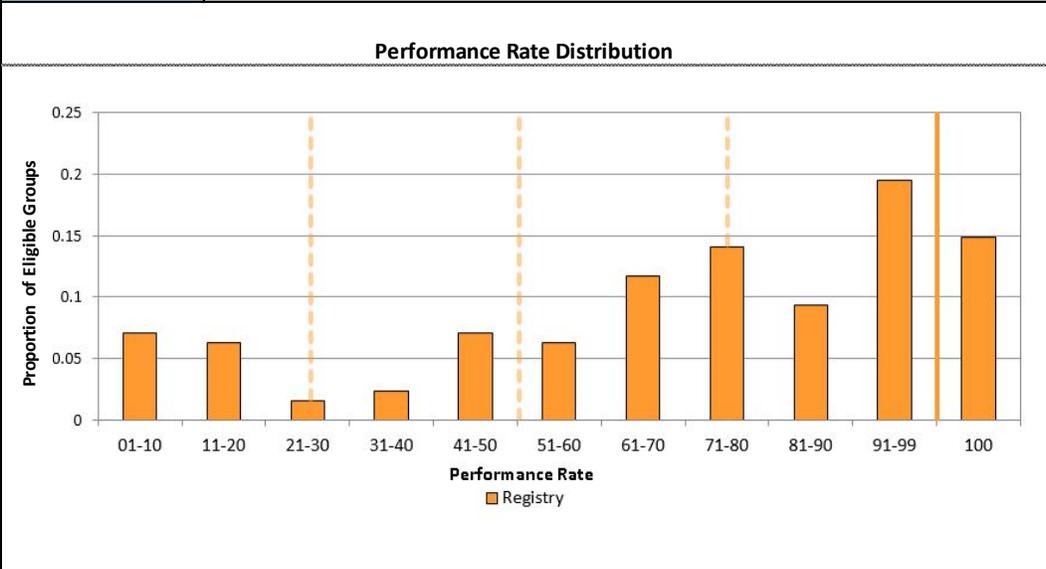


Measure	PQRS 47 - ADVANCED CARE PLAN
Description	Percentage of patients aged 65 years and older who have had an advanced care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advanced care plan was discussed but the patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan.



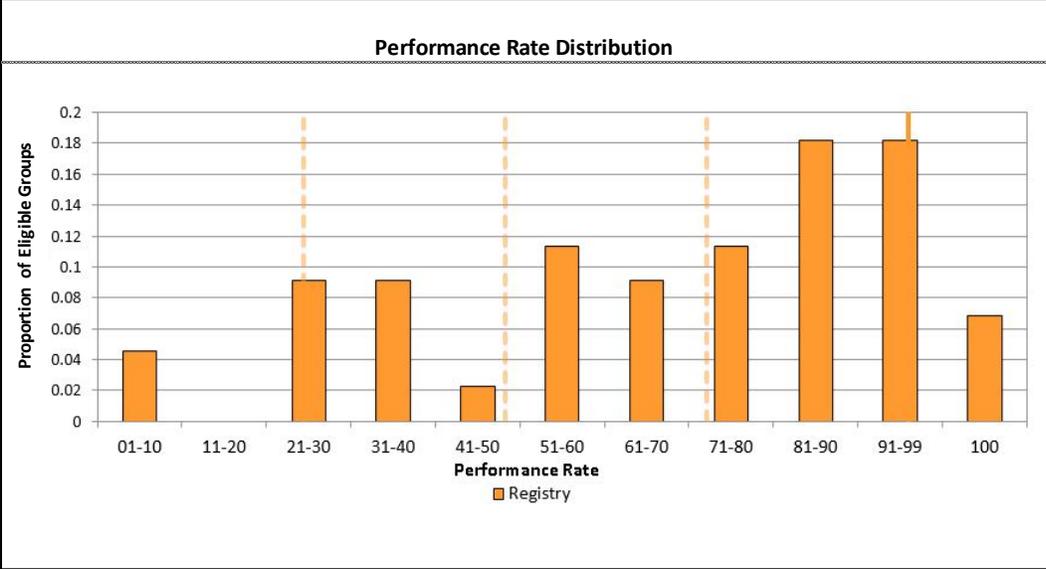
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Registry	667	79.2%	25.1%	0.1%	67.5%	89.4%	98.8%	100.0%

Measure	PQRS 48 - URINARY INCONTINENCE: ASSESSMENT OF PRESENCE OR ABSENCE OF URINARY INCONTINENCE IN WOMEN AGED 65 YEARS AND OLDER
Description	Percentage of female patients aged 65 years and older who were assessed for the presence or absence of urinary incontinence within 12 months.



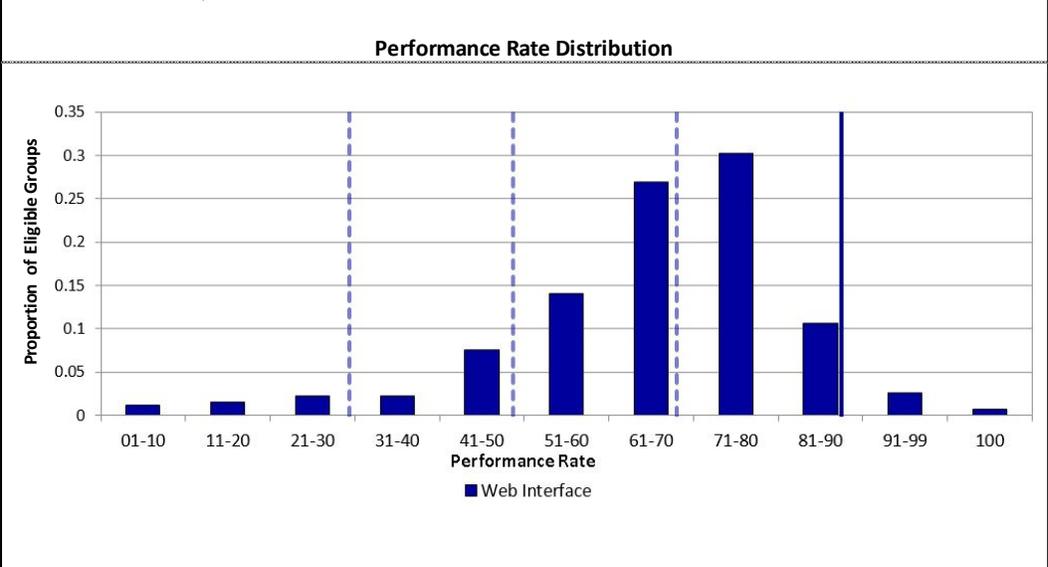
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Registry	128	69.2%	30.1%	0.5%	53.6%	77.6%	94.7%	100.0%

Measure	PQRS 51 - CHRONIC OBSTRUCTIVE PULMONARY DISORDER (COPD): SPIROMETRY EVALUATION
Description	Percentage of patients aged 18 years and older with a diagnosis of COPD who had spirometry results documented.



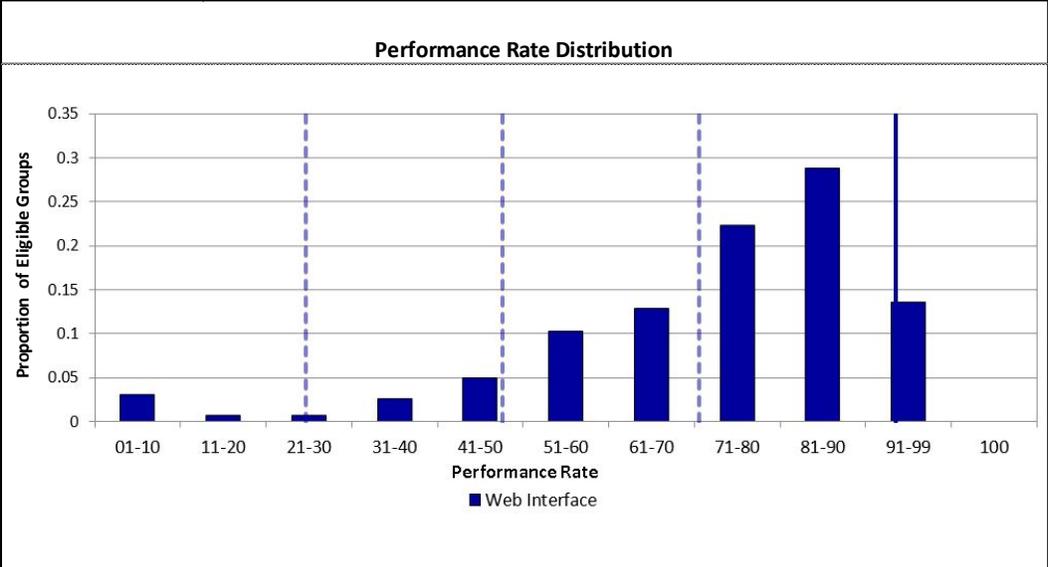
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Registry	44	67.1%	28.0%	0.6%	48.0%	75.3%	91.0%	100.0%

Measure	PQRS 110 - PREVENTIVE CARE AND SCREENING: INFLUENZA IMMUNIZATION
Description	Percentage of patients aged 6 months and older seen for a visit between October 1 and March 31 who received an influenza immunization OR who reported previous receipt of an influenza immunization.



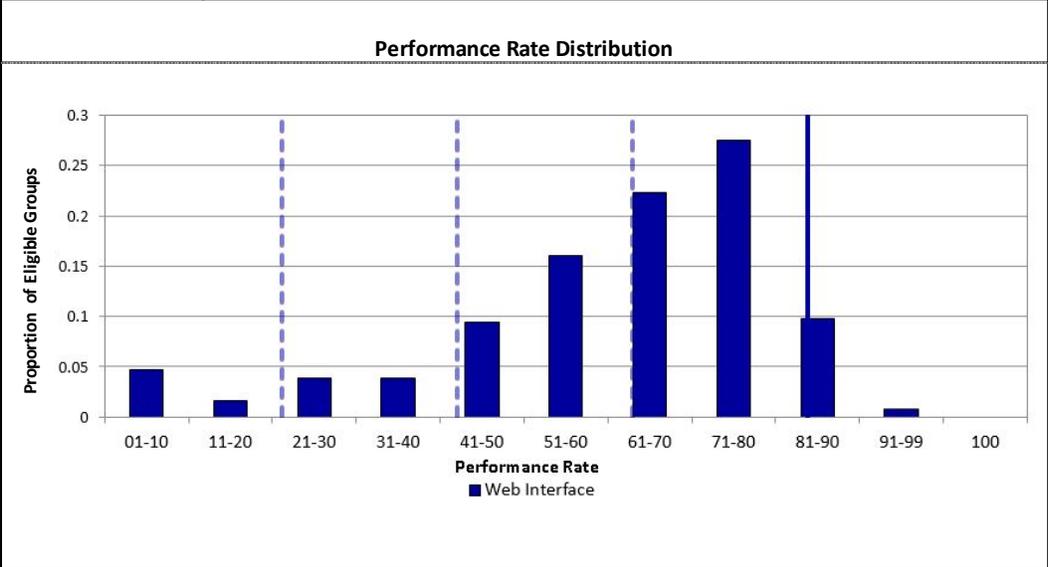
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	264	66.1%	16.6%	0.8%	58.4%	68.9%	77.0%	100.0%

Measure	PQRS 111 - PNEUMONIA VACCINATION STATUS FOR OLDER ADULTS
Description	Percentage of patients 65 years of age and older who have ever received a pneumococcal vaccine.



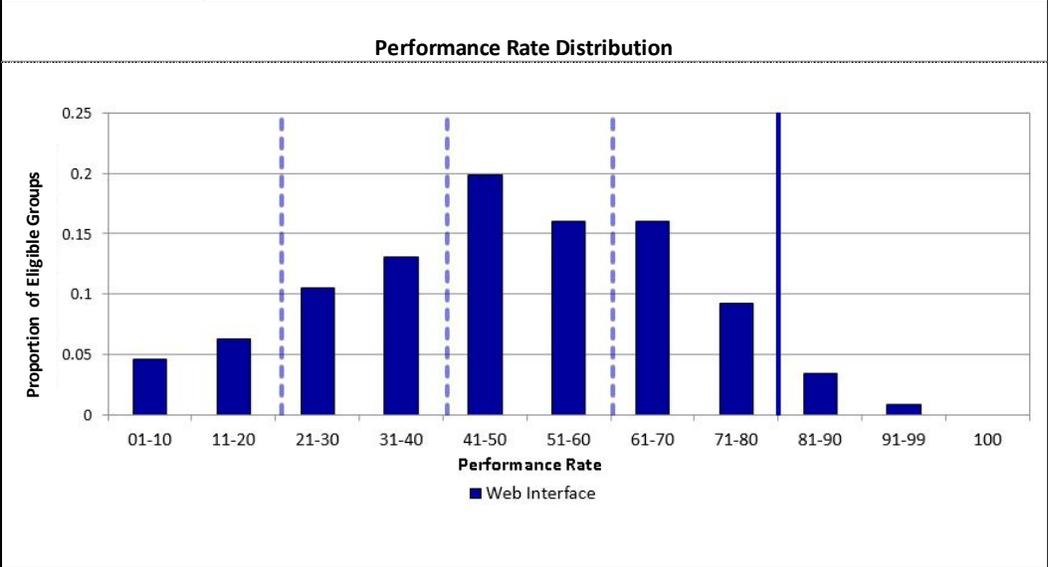
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	264	72.1%	20.4%	1.6%	63.9%	77.2%	86.4%	98.8%

Measure Description	PQRS 113 - COLORECTAL CANCER SCREENING Percentage of patients 50 through 75 years of age who had appropriate screening for colorectal cancer.
----------------------------	---



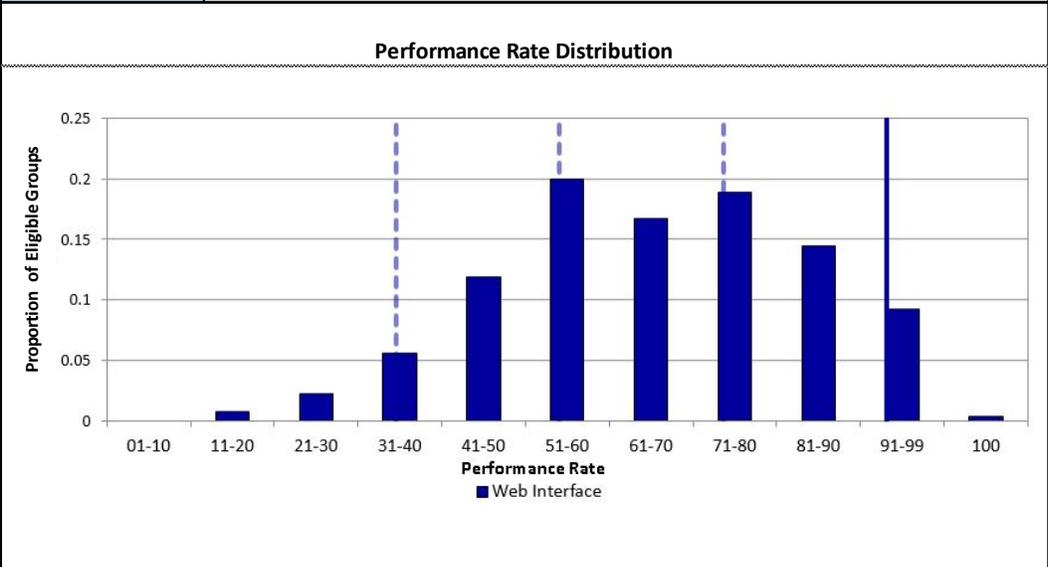
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	255	60.6%	20.4%	0.4%	51.9%	66.0%	76.1%	91.5%

Measure Description	PQRS 117 - DIABETES: EYE EXAM Percentage of patients 18 through 75 years of age with a diagnosis of diabetes (type 1 and type 2) who had a retinal or dilated eye exam by an eye care professional in the measurement period or a negative retinal or dilated eye exam (negative for retinopathy) in the year prior to the measurement period.
----------------------------	--



Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	237	47.9%	20.4%	0.9%	33.9%	48.8%	64.0%	96.0%

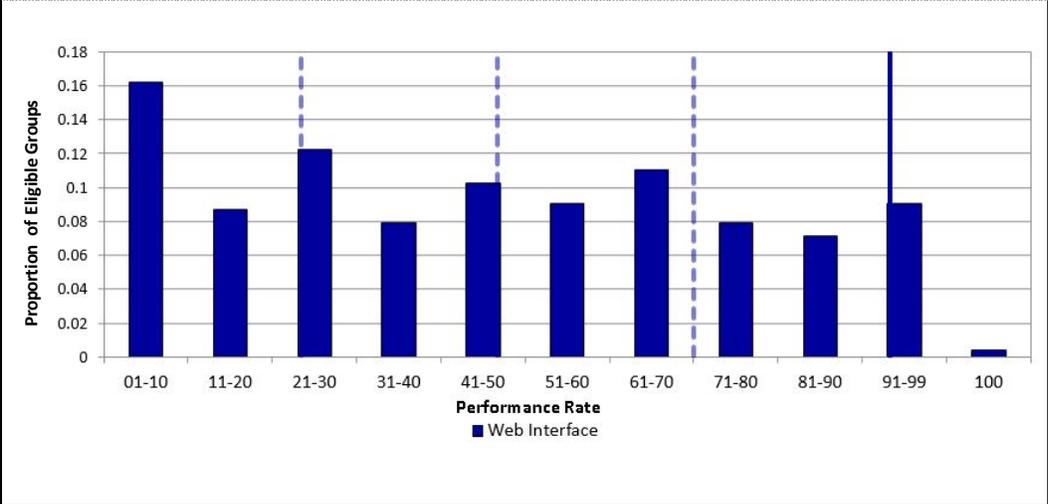
Measure	PQRS 128 - PREVENTIVE CARE AND SCREENING: BODY MASS INDEX (BMI) SCREENING AND FOLLOW-UP PLAN
Description	Percentage of patients aged 18 years and older with a BMI documented during the current encounter or during the previous six months AND with a BMI outside of normal parameters, a follow-up plan is documented during the encounter or during the previous six months of the current encounter.



Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	270	65.8%	18.3%	15.0%	52.9%	66.8%	79.8%	100.0%

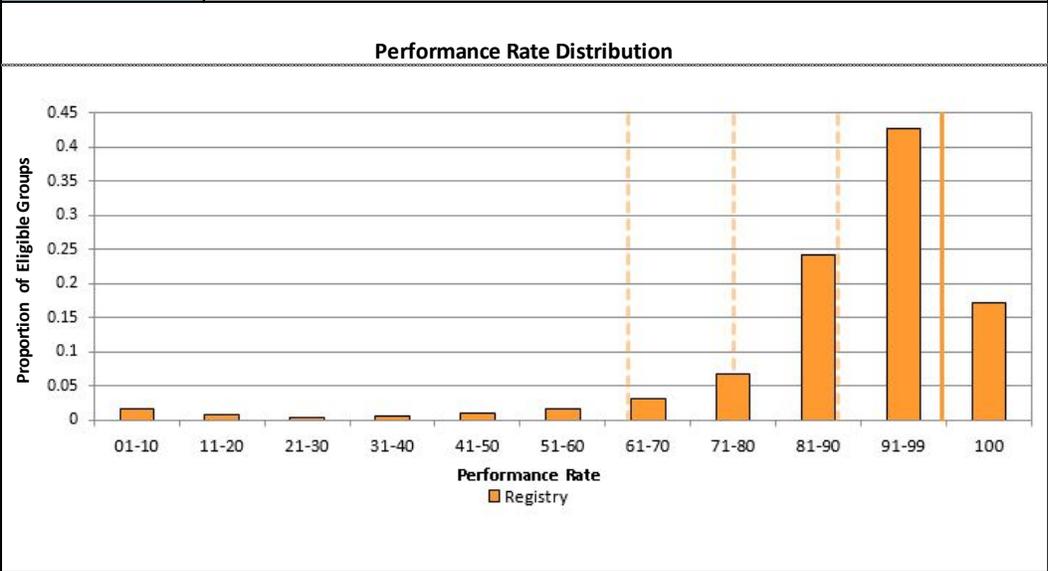
Measure	PQRS 134 - PREVENTIVE CARE AND SCREENING: SCREENING FOR CLINICAL DEPRESSION AND FOLLOW-UP PLAN
Description	Percentage of patients aged 12 years and older screened for clinical depression on the date of the encounter using an age appropriate standardized depression screening tool AND if positive, a follow-up plan is documented on the date of the positive screen.

Performance Rate Distribution



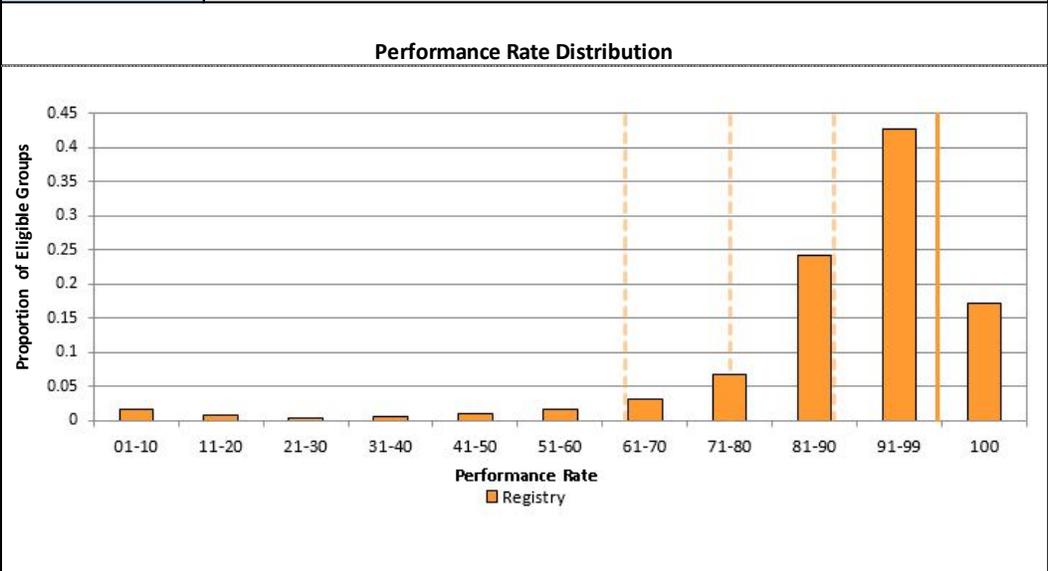
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	253	45.8%	29.4%	0.2%	20.8%	44.6%	69.4%	100.0%

Measure	PQRS 226 - PREVENTIVE CARE AND SCREENING: TOBACCO USE: SCREENING AND CESSATION INTERVENTION
Description	Percentage of patients aged 18 years old or older who were screened for tobacco use one or more times within 24 months AND who received cessation counseling intervention if identified as a tobacco user.



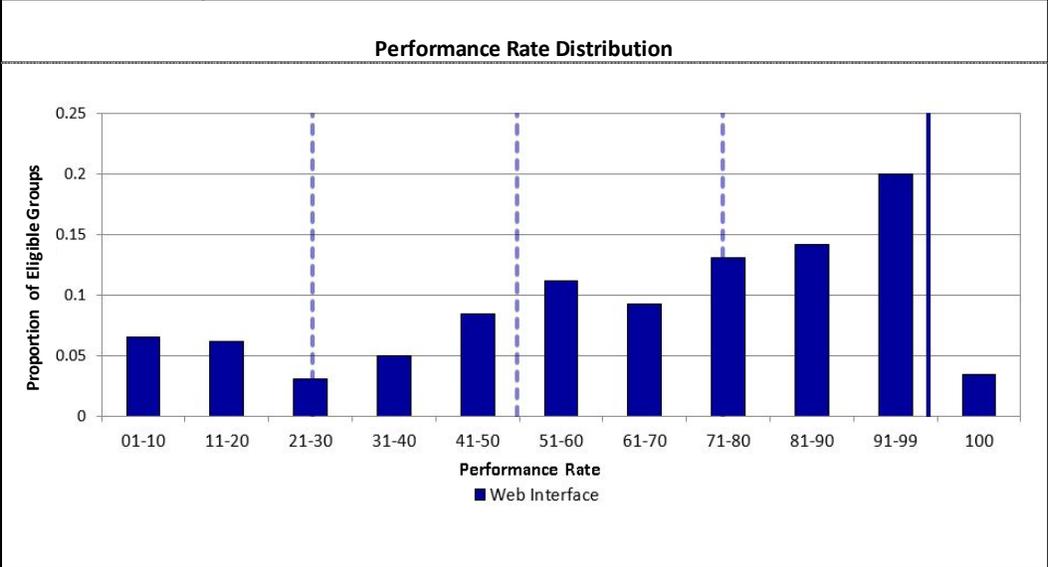
Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Registry	1,659	87.9%	17.2%	0.2%	86.1%	92.5%	98.0%	100.0%

Measure	PQRS 238 - USE OF HIGH-RISK MEDICATIONS IN THE ELDERLY
Description	Percentage of patients 66 years of age or older who were ordered at least one high-risk medication.



Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Registry	654	91.0%	9.5%	46.5%	86.1%	93.6%	99.0%	100.0%

Measure Description	PQRS 318 - FALLS: SCREENING FOR FALL RISK Percentage of patients 65 years of age and older who were screened for future fall risk at least once during the measurement period.
----------------------------	--



Reporting Mechanism	Number of Groups	Group Performance Rates						
		Mean	Std Dev	Min	25th	Median	75th	Max
Web Interface	261	64.4%	28.5%	0.6%	47.2%	71.0%	89.0%	100.0%