EXECUTIVE SUMMARY

The Centers for Medicare & Medicaid Services (CMS) has made quality improvement a central component of its efforts to transform health care, and as such has been active in expanding and strengthening quality incentive and value-based purchasing programs used across all aspects of the Medicare program. The Medicare Advantage (MA) Star Ratings program is a quality-based incentive program that links Medicare plan payments to health plan performance on a variety of quality measures. It is one of a number of quality incentive programs CMS uses in the Medicare program to link program payments to quality performance. While the other Medicare quality incentive programs impact fee-for-service (FFS) provider payments under Medicare Parts A and B, the MA Star Ratings are used to adjust the capitated rates that MA plans receive under Part C.

This paper compares the MA Star Ratings program to a number of other quality incentive programs used in FFS Medicare. In examining these programs we focused on four main areas of comparison: program processes, measure sets, scoring methodology, and transparency of quality information for consumers. We found key differences between the MA Star Ratings program and other Medicare quality incentive programs, which highlight areas that could be improved in the MA Star Ratings. Additionally, we identified issues common across multiple programs that could help inform CMS as it moves forward with refining and strengthening these programs overall.

Below we present key findings from this work and corresponding recommendations based on those findings:

Program Processes

• Unlike every other Medicare quality incentive program reviewed including those for hospitals and physicians, the MA Star Ratings program is the only one that announces and implements program changes without affording participating plans the benefits of a formal rulemaking process. As a result, unlike other programs, the MA Star Ratings program does not give stakeholders a full 60-day comment opportunity on program changes such as new measures or score calculation methodology updates, despite significant plan payment implications.

• The MA Star Ratings is the only quality incentive program whose measure set is not finalized before the performance period, meaning that unlike hospitals and other providers, plans do not know the finalized measure set before the data are collected. The other quality incentive programs finalize the measures to be included in the measure set before the actual data collection period for the measures. In contrast, the MA Star Ratings measure set for a given year is finalized during or after the data collection period for the measures.

Recommendation on Program Processes:

1. CMS should implement and use a formal rulemaking process for the MA Star Ratings program that provides a 60-day comment opportunity, has earlier and more thorough vetting of new measures, and finalizes measure sets and scoring methodology in advance of the data collection and performance periods.

Measure Sets

• MA plans are evaluated on a greater number of measures, in total, than other groups, including Accountable Care Organizations (ACOs). Currently MA plans that offer Part D (MA-PDs) are evaluated on 44 quality measures—more measures than any of the other value-based purchasing programs. The performance measure set for Accountable Care Organizations (ACOs) includes just 31 measures in 2017.
• Across most programs reviewed, outcome measures remain underrepresented, despite efforts to value these measures by weighting them more. Although CMS has expressed an intention to emphasize quality measures that focus on health outcomes—as opposed to processes—only 20 percent of the current measures used for MA-PD ratings are outcome or intermediate outcome measures, and even with greater weighting than process measures, they account for only 33 percent of the overall scores for MA-PDs.

• Measures across many programs continue to need better adjustment for socioeconomic status of beneficiaries served. Measure scores are often not adjusted, or are insufficiently adjusted, to account for the socioeconomic differences of beneficiaries that plans and providers may serve. Recent research suggests this disadvantages plans or providers that serve a disproportionate share of beneficiaries with lower socioeconomic status.

**Recommendations on Measure Sets:**

2. CMS should reduce the number of measures included in the MA Star Ratings, while focusing on clinically meaningful outcome measures to drive greater improvements in health outcomes.

3. CMS should continue to examine and test more meaningful adjustments to plans’ ratings to account for differences in the socioeconomic status of enrollees.

**Scoring Methodology**

• Unlike in other programs where the measure of year-to-year improvement can only increase scores, the improvement measures in the MA Star Ratings program can penalize plans. The MA Star Ratings improvement measures can have negative implications for plans, even under the policy to hold high performing plans harmless. This is unique among the programs reviewed. Further, not all plans are able to benefit from the improvement measures, as these measures are not included in the score calculations for lower performing plans.

• CMS recognizes the value of performance thresholds to help drive improvement in other programs, but has recently stopped using them in the MA Star Ratings. CMS will use performance thresholds to incentivize physician improvements under the recently finalized Medicare Access and CHIP Reauthorization Act (MACRA) rules. While the agency acknowledges the importance of providing goals to help focus improvements, CMS recently eliminated the thresholds for MA plans despite plans’ objections; many MA plans had used the thresholds to incentivize performance among network providers.

• For a number of programs examined, computing a total quality score is complex and can include multiple steps. For instance, scoring methodologies assign relative values to performance levels, combine measures and/or measure categories, apply weighting methodologies, and incorporate adjustments. These factors create multiple layers and moving parts for plans and providers to track, diverting resources away from improvement efforts. Further, the complicated methodologies can divorce actual performance on a measure from overall scores.
Opportunities to Strengthen the Medicare Advantage Star Ratings Program:
A Comparison to Other Medicare Quality Incentive Programs
Public Policy Institute

Consumer Information

- The MA Star Ratings offer a more consumer friendly model for presenting quality information to beneficiaries, but there is no way to compare MA plan quality to traditional FFS Medicare quality. The five star rating system for the MA Star Ratings program and its display on the Plan Compare website are much less complicated than the information presented on Hospital Compare or Physician Compare. However, there is not sufficient comparable information about the quality of care in the FFS Medicare program for beneficiaries to be able to compare plan quality to FFS in making enrollment decisions.

Recommendation on Consumer Information:

7. To aid Medicare beneficiaries in making a fully informed enrollment choice that is based on both cost and quality considerations, CMS should provide comparable measures of FFS quality for beneficiaries to consider alongside MA plans’ Star Ratings.

Conclusion

As CMS continues to explore ways to strengthen quality improvement programs in Medicare, it can build off the experiences it has in implementing multiple types of programs for various plans and providers. These recommendations are steps that CMS can take to continue to strengthen how the MA Star Ratings program works and to better drive quality improvements for beneficiaries.

Recommendations on Scoring Methodology:

4. Given the complexity and multitude of quality incentive programs that providers face, CMS should continue to work towards greater alignment of measures and objectives between FFS and MA plans.

5. CMS should review the methodology used for the improvement measures in Part C and Part D to ensure it rewards plans for improvements in performance, protects plans that are higher performing from being penalized, and promotes improvement opportunities for all plans including lower performing plans.

6. CMS should reinstate four-star thresholds for selected measures to more clearly identify desired levels of performance based on prior year trends, as well as expected and realistic improvement standards, and to encourage stable improvement objectives.
INTRODUCTION

Quality Incentive Programs in Medicare

There has been a growing emphasis among policymakers, government agencies, employers, and health plans on paying for health care based on the quality of care delivered rather than only on the quantity of services provided. There are many types of quality incentive programs, which can include things such as pay-for-reporting or pay-for-performance programs, bonus payments, payment penalties and value-based purchasing (VBP) programs. These programs are intended to drive spending in health care toward higher value and more efficient providers and services. In particular, policymakers have focused on expanding the use and reach of pay-for-performance programs in Medicare, especially following enactment of the Affordable Care Act (ACA), which included a number of provisions that introduced new quality incentive programs or strengthened existing ones.1

One of the quality incentive programs enacted under the ACA is the Medicare Advantage (MA) Star Ratings (“MA Star Ratings”) program which links MA plans’ payments to their performance on quality measures. The MA Star Ratings focus on measuring the quality of private plans providing the Medicare benefit to enrollees under Medicare Part C, and usually Part D as well. There are also a number of quality incentive programs used in fee-for-service (FFS) Medicare applicable to providers such as hospitals, physicians, home health agencies (HHAs), and dialysis centers, to name a few. The FFS programs range in type from pay-for-reporting of quality measures to more robust VBP programs. We examined the MA Star Ratings program in the context of a number of other Medicare FFS quality incentive programs as shown in Table 1 (see Appendix A for an overview of Quality Incentive Programs discussed in this paper).

Table 1. Medicare Quality Incentive Programs Reviewed in This Study

<table>
<thead>
<tr>
<th>Provider/Plan Type</th>
<th>Quality Incentive Programs Reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicare Advantage</td>
<td>• MA Star Ratings (including Part D measures)</td>
</tr>
<tr>
<td>Hospitals*</td>
<td>• Hospital Inpatient Quality Reporting (IQR) and Outpatient Quality Reporting (OQR)</td>
</tr>
<tr>
<td></td>
<td>• Hospital Readmissions Reduction Program (HRRP)</td>
</tr>
<tr>
<td></td>
<td>• Hospital Acquired Condition (HAC) Reduction Program</td>
</tr>
<tr>
<td></td>
<td>• Hospital Value Based Purchasing (VBP)</td>
</tr>
<tr>
<td>Accountable Care Organizations (ACOs)</td>
<td>• ACO Quality Measure Program</td>
</tr>
<tr>
<td>Physicians</td>
<td>• Physician Quality Reporting System (PQRS)</td>
</tr>
<tr>
<td></td>
<td>• Meaningful Use</td>
</tr>
<tr>
<td></td>
<td>• Physician Value-Based Payment Modifier (VBPM)/Merit-Based Incentive Payment System (MIPS)</td>
</tr>
<tr>
<td>Dialysis Facilities</td>
<td>• End Stage Renal Disease (ESRD) Quality Improvement Program (QIP)</td>
</tr>
<tr>
<td>Home Health Agencies</td>
<td>• Home Health Agency (HHA) Quality Reporting</td>
</tr>
<tr>
<td></td>
<td>• HHA Pilot VBP</td>
</tr>
<tr>
<td>Skilled Nursing Facilities (SNFs)</td>
<td>• Skilled Nursing Facilities (SNF) VBP</td>
</tr>
</tbody>
</table>

*There are additional quality incentive programs which may impact hospital payments, such as for inpatient psychology units, which were not included in this paper.
Quality incentive programs can be structured in a variety of ways. The programs examined for this report fall into one of five types: pay-for-reporting, payment penalties, bonus payments, VBP programs, or shared savings eligibility (see Table 2). In some instances more than one type of incentive program may apply to a particular provider group. For example, inpatient hospitals are subject to pay-for-reporting requirements (Inpatient Quality Reporting (IQR)), payment penalties (Hospital Readmissions Reduction Program (HRRP) and Hospital Acquired Conditions (HAC) Program), as well as a VBP program (Hospital VBP). A number of the pay-for-performance programs or value-based purchasing programs evolved from longer-standing pay-for-reporting requirements. CMS, consistent with the direction in the ACA and the Medicare Access and CHIP Reauthorization Act (MACRA), is moving away from simply incentivizing reporting to using quality information to tie payment more closely to higher quality performance.

MA plans receive a bonus to their base payment levels for higher quality performance levels. Despite a mostly upside system for MA plans in terms of quality performance, it should be noted that there is also an inherent downside from a competitive perspective (such as less money for extra benefits to attract beneficiaries). MA plans that consistently perform poorly on the MA Star Ratings can be removed from program participation, so there is a very real negative consequence for poor quality performance by plans.

Table 2. Types of Quality Incentive Programs in Medicare

<table>
<thead>
<tr>
<th>Type of Program</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pay-for-Reporting</td>
<td>A portion of payments are withheld unless providers report quality measures or fulfill other programmatic requirements</td>
<td>• PQRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• IQR and OQR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• HHA Quality Reporting</td>
</tr>
<tr>
<td>Payment Penalties</td>
<td>Payments are reduced for covered entities whose quality performance on certain measures falls below specified thresholds or is low relative to the group</td>
<td>• HRRP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• HAC Reduction Program</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• ESRD QIP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Meaningful Use</td>
</tr>
<tr>
<td>Bonus Payments</td>
<td>Covered entities are rewarded through greater payments for achieving higher quality performance</td>
<td>• MA Star Ratings</td>
</tr>
<tr>
<td>VBP Programs</td>
<td>Payments may be increased, decreased, or unchanged based on quality performance. VBP programs withhold a portion of payments from all entities and then redistribute the withheld payments based on performance</td>
<td>• Hospital VBP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Physician VBPM/MIPS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• SNF VBP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• HHA Pilot VBP</td>
</tr>
<tr>
<td>Shared Savings Requirement</td>
<td>Organizations are not able to share in savings achieved through better care management if they do not meet quality thresholds</td>
<td>• ACO Quality Measure Program</td>
</tr>
</tbody>
</table>

Payment Implications

In general, the quality incentive programs reviewed affect 1 percent to 5 percent of payments for any given provider (see Appendix A for program specific details). Some programs like the Hospital VBP and the Physician VBPM had payment implications phased in to their eventual levels, which are 2 percent for the Hospital VBP in 2017 and thereafter and 2 percent to 4 percent for the Physician VBPM in 2017, depending on the physician practice group size (beginning in 2019 physicians will be paid under the Merit-Based Incentive Payment System (MIPS)). As discussed above, it is important to note that some provider groups are affected by multiple programs and therefore have a greater total share of their payments impacted by quality programs. For example, 6 percent of inpatient
hospital payments are tied to performance, which does not include the penalties associated with not reporting on required quality measures. Quality scores also impact ACO payments, preventing shared savings payments if the ACO does not report the measures or perform at sufficient quality levels. Finally, as described below in more detail, MA plans that achieve 4-5 Stars see significant revenue boost—their payment benchmarks are increased by 5 percent to 10 percent of county level FFS costs.

Overview of the MA Star Ratings

In the MA program, private health plans provide the Medicare benefit to enrollees as an alternative to traditional Medicare, or Medicare FFS. Although plans are required to cover all Medicare services, they have some flexibility in how they design their benefits, they may limit choice of providers to those in a plan’s network, and they can offer additional benefits beyond FFS. The ACA required significant changes to MA plan payments, including linking MA plan payments to the plan’s quality rating, or MA Star Rating.

MA Plan Payments Are Tied to MA Star Ratings

MA plans are paid based on administratively set county-specific payment rates (or “benchmarks”). The ACA aligned the MA county level benchmarks more closely with FFS costs in each county. Plans submit bids representing their estimated costs for providing the Medicare Parts A and B benefits and the MA plan bids are compared to the benchmarks. If a plan bids below the benchmark, it retains a portion of the difference between its bid and the benchmark, which is known as a “rebate.” Rebates can be used to provide extra benefits for enrollees such as reduced cost sharing or enhancing the Part D benefit if the plan offers prescription drug coverage. If a plan bids above the benchmark, it must charge a premium to the enrollee for the amount above the benchmark. Under the ACA, higher performing plans—those with 4 or more Stars—are eligible for higher county benchmarks, and plans with 3.5 Stars or more are eligible for larger rebates.

The MA Star Ratings use a five-star quality rating system to measure MA plan performance. As mentioned above, there are two ways plans are rewarded based on their Star Ratings. The first is through increases to the benchmarks. In most counties, plans with 4-5 Stars receive a bonus in the form of an increase to their county benchmark equal to 5 percent of the county’s underlying FFS costs (meaning plans bid against a higher amount and are therefore eligible for larger rebates if they bid below the benchmark or do not have to charge as large a premium if they bid above the benchmark). In some counties, known as double bonus counties, the benchmark is increased by 10 percent of underlying FFS costs. However, because the ACA limited benchmarks to no more than they would have been prior to the ACA, in some counties, this “benchmark cap” prevents plans from receiving the full amount of the quality bonus they earned.

Second, MA Star Ratings determine the rebates that plans receive when they bid below the benchmarks. Prior to the ACA, MA plans received a rebate equal to 75 percent of the difference between the plan’s bid and the benchmark. The ACA created a tiered system for rebate amounts based on the Star Ratings as shown below, with the highest performing plans retaining the largest portion of the difference.

Table 3. MA Performance Based Payment Changes

<table>
<thead>
<tr>
<th>Plan Star Rating*</th>
<th>Benchmark Increase</th>
<th>Rebate Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 to 5 Stars</td>
<td>5% of county FFS (10% in double bonus counties)</td>
<td>70%</td>
</tr>
<tr>
<td>4 Stars</td>
<td>5% of county FFS (10% in double bonus counties)</td>
<td>65%</td>
</tr>
<tr>
<td>3.5 Stars</td>
<td>N/A</td>
<td>65%</td>
</tr>
<tr>
<td>3 or Less Stars</td>
<td>N/A</td>
<td>50%</td>
</tr>
</tbody>
</table>

*New plans from new parent organizations receive a benchmark increase equal to 3.5% of county FFS costs and 65% rebate amount.
Determination of the MA Star Ratings

MA plans are evaluated on a variety of measures, each of which is also rated on a five-star scale. Plans then receive an overall score based on weighted performance across the measures. For MA plans that provide Part D coverage (also known as MA-PDs), the overall score includes performance on both Part C and Part D measures. For 2017, there are 32 Part C measures and 15 Part D measures. MA-PD plans’ overall scores reflect performance on 44 measures; three measures overlap and are included in both the MA and Part D measure sets but are included only once in the overall MA-PD score.

The MA Star Ratings include different categories of measures. These categories are process measures, beneficiary experience measures, outcome and intermediate outcome measures, access measures and improvement measures. The measures are differentially weighted in the score calculation to emphasize the importance of some measures (see Table 4). For example, outcome measures have three times the contribution to the overall score as process measures. The improvement measures—used to evaluate year-to-year changes in performance—have the greatest weight (5 times the process measures).

In order to determine the star rating for each of the individual measures, CMS first compares the “raw performance” across all plans on each measure for a year and looks for statistically significant “cutpoints.” These cutpoints determine the scores needed to achieve each level of stars (e.g., 3 Stars, 4 Stars). In this way, for most measures, MA plans are rated on a performance curve and the actual scores needed to achieve a particular Star Rating can change annually. Once a plan’s measure-specific

<table>
<thead>
<tr>
<th>Measure Category</th>
<th>Weight in Summary and Overall Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Beneficiary Access or Beneficiary Experience and Complaints</td>
<td>1.5</td>
</tr>
<tr>
<td>Intermediate Outcome or Outcome</td>
<td>3</td>
</tr>
<tr>
<td>Improvement</td>
<td>5</td>
</tr>
</tbody>
</table>

Star Ratings are determined, CMS calculates the improvement measure scores. Inclusion of improvement measures depend on the plan’s scores without the improvement measures in part to hold higher performing plans harmless for minor changes in performance. Finally, CMS calculates the overall score, using the measure weighting approach described above.

Following the overall score calculation, CMS makes two additional adjustments to plans’ scores. First, CMS will apply a reward factor for plans that demonstrate consistently high performance across a number of measures. The range of factors applied for high performance is 0.1 to 0.4 stars. This adjustment can have a meaningful impact on the final MA Star Rating.

Finally, beginning in 2016, CMS started to employ a categorical adjustment index (CAI) to adjust plans’ MA Star Ratings to account for the effects of serving higher proportions of low income or disabled beneficiaries. However, the actual rating adjustment amounts for the contracts with higher proportions of beneficiaries with low incomes or disabilities has been very small and for many contracts is not enough to move their overall scores. In 2017, only 15 plan contracts (or 4 percent of plan contracts) saw a change to their star ratings because of the CAI and the remaining 349 plan contracts (96 percent) saw no change. Therefore, despite the significant complexity and resource intensity of this approach, it is expected to have relatively little impact—changing the scores of only a small proportion of plans.

The MA Star Ratings affect payments nearly a year and a half after they are posted to CMS’ Plan Compare website as a tool for beneficiaries shopping for coverage. So, for example, the 2017 Star Ratings, which are published on Plan Compare in October 2016 for the 2017 annual enrollment period, are based on performance data collected in 2015 and 2016 and are then used to adjust 2018 payments, as Figure 1 shows. CMS proposes and finalizes the Star Ratings metrics and methodology in 2015-2016, while data are being collected.

Table 4. MA Star Ratings Measure Categories

The MA Star Ratings Program: A Comparison to Other Medicare Quality Incentive Programs
MA Plan Quality Is Improving

In MA, data from CMS on the distribution of contracts’ Star Ratings from 2013 to 2017 illustrates gains in quality among MA plans. In 2017, nearly 49 percent of MA-PD contracts are 4-5 Star contracts, 1.75 times higher than in 2013 when only 28 percent of contracts had 4-5 Stars. Additionally, there are only 12 contracts that are below 3 Stars in 2017, compared to 62 in 2013. Perhaps more importantly, in 2017, approximately 68 percent of MA-PD enrollment was in contracts with 4-5 Stars compared to 2013, when only 38 percent of enrollees were in higher performing plans (see Figure 2). As plans have improved their performance, it appears that enrollees are also increasingly selecting higher quality MA plans.

Figure 2. Distribution of MA-PD Enrollees by Contract Quality Score, 2013-2017

- 0% - 2 Stars
- 5% - 2.5 Stars
- 10% - 3 Stars
- 15% - 3.5 Stars
- 20% - 4 Stars
- 25% - 4.5 Stars
- 30% - 5 Stars

2013 - Blue
2014 - Purple
2015 - Green
2016 - Orange
2017 - Brown

Opportunities to Strengthen the Medicare Advantage Star Ratings Program: A Comparison to Other Medicare Quality Incentive Programs
Opportunities to Strengthen the Medicare Advantage Star Ratings Program:  
A Comparison to Other Medicare Quality Incentive Programs

Public Policy Institute

STUDY FINDINGS: COMPARING MEDICARE QUALITY INCENTIVE PROGRAMS

Program Processes

MA Star Ratings Is the Only Program that Does Not Announce Program and Measure Changes through a Rulemaking Process

In order to introduce and implement changes to many of the programs under its authority, CMS goes through a rulemaking process which includes both notice of proposed changes and an opportunity for stakeholders to respond and comment to the proposed changes. Of all the programs examined, the MA Star Ratings program is the only one that does not follow a formal rulemaking process for announcing annual program changes and updates (see details in Table 5). Every other Medicare quality incentive program examined, including those that only have reporting requirements, goes through a formal rulemaking process including a 60-day comment period for stakeholders. For the MA Star Ratings, CMS uses a sub-regulatory guidance process with shorter comment opportunities.

Table 5. Announcement of Annual Changes to Medicare Quality Incentive Programs

<table>
<thead>
<tr>
<th>Programs</th>
<th>Announcement Process</th>
<th>Comment Opportunity</th>
<th>Rule or Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA Star Ratings</td>
<td>Guidance and Sub-regulatory Guidance with Comment Opportunity</td>
<td>2 weeks following the November Preview and 30 days following the February Draft Call Letter (if continues to be released with Advance Notice)⁹</td>
<td>November Preview &amp; February Draft Call Letter/ April Final Call Letter</td>
</tr>
<tr>
<td>ACOs</td>
<td>Notice of Proposed Rule-making (NPRM) with Comment Opportunity</td>
<td>60 Days</td>
<td>Medicare Shared Savings Program Rule</td>
</tr>
<tr>
<td>Hospital IQR, HAC, HRRP, VBP,</td>
<td>NPRM with Comment Opportunity</td>
<td>60 Days</td>
<td>Inpatient Prospective Payment System (IPPS)</td>
</tr>
<tr>
<td>Hospital OQR</td>
<td>NPRM with Comment Opportunity</td>
<td>60 Days</td>
<td>Outpatient Prospective Payment System (OPPS)</td>
</tr>
<tr>
<td>PQRS, Physician VBPM, MIPS</td>
<td>NPRM with Comment Opportunity</td>
<td>60 Days</td>
<td>Physician Fee Schedule</td>
</tr>
<tr>
<td>ESRD QIP</td>
<td>NPRM with Comment Opportunity</td>
<td>60 Days</td>
<td>ESRD Prospective Payment System (PPS)</td>
</tr>
<tr>
<td>HHA Quality Reporting, HHA VBP Pilot Program</td>
<td>NPRM with Comment Opportunity</td>
<td>60 Days</td>
<td>HHA PPS</td>
</tr>
<tr>
<td>SNF VBP</td>
<td>NPRM with Comment Opportunity</td>
<td>60 Days</td>
<td>SNF PPS</td>
</tr>
</tbody>
</table>
One reason for this difference is that CMS generally includes information about incentive programs in the annual rules announcing payment updates for a provider type (e.g., inpatient hospitals) since the incentive programs have payment implications. So, for example, all of the changes to the measures and methodologies associated with inpatient hospital quality incentive programs like HRRP, HAC, or the inpatient hospital VBP are all announced through the annual inpatient prospective payment system (IPPS) rulemaking. Similarly, the physician quality incentive programs are announced through the physician fee schedule proposed and final rules. Since MA payment changes are announced through the Advance Notice and Draft Call Letter (ANCL), which is sub-regulatory guidance, CMS also uses the Call Letter to announce changes to the MA Star Ratings program. However, this results in a significant difference in opportunity to evaluate and comment on proposed changes. Additionally, CMS is not required to consider and respond to each comment as it must do with proposed regulations.

However, plans have raised concerns that given the payment implications and the complexity of the programs, CMS should implement a more formal process for announcing changes to the MA Star Ratings or, at a minimum, provide a longer comment opportunity.

To address these concerns in part, in 2013 CMS began previewing MA Star Ratings changes in November through a Request for Comments (RFC) with a two week comment opportunity. The language from the preview is often repeated verbatim in the Draft Call Letter that is then released in February of the next year. Historically, there was a two week comment opportunity following that release. Recent legislative changes seem to provide additional time for review as plans now have 30 days to comment on the Advance Notice (and if the Draft Call Letter continues to be released concurrently with the Advance Notice then the longer comment opportunity for proposed changes to the MA Star Ratings will apply). Still, plans do not get the 60-day comment opportunity associated with the formal rulemaking process for other Medicare providers.

Further, it seems the lengthening of the February comment period prompted CMS to reduce the information provided in the November preview. For the 2018 preview, CMS deviated from the process of providing the full list of proposed changes to the Star Ratings in the November release and only focused on requesting comments on proposed changes to one specific measure. In the Draft Call Letter for 2018, a number of other changes to the Star Ratings were included for comment.

New MA Star Ratings Measures Are Introduced and Incorporated Differently than Are Measures in Other Programs

The ACA established a vetting process for new quality measures included in most of the FFS quality incentive programs, known as the "pre-rulemaking process" which announced and vetted Measures under Consideration (MUC). This process includes an opportunity for stakeholders to comment on the measures before they are selected for use. Additionally, when the MUC are released for comment, CMS provides detailed specifications and rationales for the measures for stakeholders to review. It is important to note that during this process, attention is given to balancing interests among stakeholders and there is significant participation from consumers and purchasers. However, the MA Star Ratings measures do not undergo this same vetting process before being included in the program.

Notice of Measure Changes and Measures under Consideration

The MUC process annually affords stakeholders, including groups like the National Quality Forum (through the Measure Applications Partnership (MAP)), to comment on various proposed measures. CMS considers the input received through this process before even proposing measures for use in various FFS quality incentive programs through formal rulemaking. While there are requirements for which programs must use the MUC process, CMS conducts the MUC process for every program examined in this paper except the MA Star Ratings program.

The MUC process creates a multi-year timeline for announcing new measures or measures that are under consideration for inclusion, giving stakeholders ample opportunities for comment. As an example, during the process of vetting new measures for the Hospital IQR, CMS will first announce for comment that a measure is under consideration nearly a year and a half before introducing it as a proposed measure for collection in the IPPS proposed rule. And, as discussed in more detail below, CMS will finalize inclusion of the new measure for reporting before hospitals are required to report on it. Further, there are additional steps that CMS must take for new measures to be included in the Hospital VBP program, which as a prerequisite requires...
that the measures be part of the Hospital IQR. In addition, measures used in the VBP program must also have been posted on Hospital Compare for at least one year and all new measures for the VBP must be announced at least 60 days prior to the beginning of the performance period (the period when data will be collected for the measure).

The MUC process also provides stakeholders with detailed information about the measures that are under consideration. Not only does CMS provide a description of each measure, but CMS also provides the measure specifications and rationales for possible inclusion in the various programs. The measure specification is a technical description of the numerator and denominator of a measure, including any pertinent inclusion or exclusion criteria, which aids stakeholders in evaluating the measure’s usefulness. The measure rationales are “the peer-reviewed evidence justifying the measure, and/or the impact the measure is anticipated to achieve.”

However, it is important to note that the MUC process does have issues. First, while it affords stakeholders more transparency into the measure selection process, many of the measures that CMS previews through the MUC process do not eventually get incorporated in the quality programs. Further, stakeholders including the American Hospital Association have raised concerns that not all of the measures are sufficiently tested or NQF endorsed, or that they may be overly focused on measuring processes unrelated to health outcomes. Finally, although CMS collects input on the proposed measures, they are still able to select measures for programs that may not have received sufficient support for inclusion through the MUC process.

**MA Star Ratings Measure Selection**

For the MA Star Ratings, the process of introducing new measures is less prescribed (though CMS does try to announce new measures at least two years in advance of collecting performance data for them). Additionally, CMS often tests new measures for inclusion in the MA Star Ratings by having plans report data and posting the results to the “Display Page.” CMS uses what it calls the Display Page to collect quality data or test quality measures that are not included in the MA Star Ratings program. The Display Page data is publicly available even if the measures do not contribute to the MA Star Ratings scores.

However, a significant difference between the MA Star Ratings measure vetting process and the MUC process is the amount of information that CMS releases about new measures, even in the Draft Call Letter. CMS only previews a description of the measure and does not provide measure specifications for stakeholders that wish to more closely examine the impact of adding the measure. Further, CMS does not regularly indicate what type of measure it is and how it will eventually be weighted if included in the MA Star Ratings. In order to better align the vetting process for the MA Star Ratings with the other quality programs, a first step would be to include a more formal announcement and vetting process with a 60 day opportunity to comment on proposed changes.

### Table 6. Measure Vetting and Measure Announcement Timing for Medicare Quality Programs

<table>
<thead>
<tr>
<th>Programs</th>
<th>Are Measures Vetted Through MUC Process?</th>
<th>Are Measures Finalized Before Performance Period?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MA Star Ratings</strong></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ACO Quality Measure Program</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hospital IQR, HAC, HRRP, VBP,</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hospital OQR</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PQRS, Physician VBPM, MIPS</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ESRD QIP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HHA QR, HHA VBP Pilot Program</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SNF VBP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
While CMS has usually done the “preview” of upcoming measures that may be included in the MA Star Ratings, they do not actually go through the pre-rulemaking process “for the selection of quality and efficiency measures” that the rest of the Medicare quality programs use. CMS does work closely with quality development, testing, and endorsement organizations in determining and selecting measures for inclusion in the MA Star Ratings. However, this neither allows plans or many other stakeholders sufficient input on the measures selected nor does it provide an adequate timeline for preparation for new measures. Further, CMS has not committed to the process of introducing new measures on the Display Page first for a period of two years before proposing to include them in the MA Star Ratings.

**In Contrast to Other Programs, the MA Star Ratings Measure Set Is Not Finalized Before the Measure Performance Period**

All other Medicare quality incentive programs finalize the measure sets and program specifications through the rulemaking process before the performance period—the period when data is collected for the measures (see Table 6). This gives providers greater certainty about what they will be measured on, before being measured, and sets clear standards and expectations for performance. Importantly, this timing also gives providers the opportunity to make investments in performance improvement prior to their having their performance assessed by the measures. This is not the case for the MA Star Ratings. The MA Star Ratings measures and processes are finalized after the performance period, creating uncertainty for plans in terms of what will be measured and exactly how performance will be evaluated. And unlike for other providers, MA plans have no ability to influence their performance on the measures in the first year(s) of measurement.

As can be seen in Figure 3, performance data for MA plans’ 2017 Star Ratings are collected in 2015 and 2016, yet the measures aren’t finalized until April of 2016. While plans’ performance data is being collected, CMS is still taking comments on proposed changes to measures, measure specifications, and score calculation methodologies. Given the timing, plans do not always have a chance to assess the implications of these proposals before the data is collected.

**Figure 3. 2017 Measure Set Announcement and Use for MA Star Ratings Compared to Hospital IQR**

<table>
<thead>
<tr>
<th>November 2014</th>
<th>CMS previews potential changes to 2017 Star Ratings in RFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2015-March 2016</td>
<td>Plans submit quality data/CMS collects data for 2017 Ratings (“Data Collection/Performance Period”)</td>
</tr>
<tr>
<td>November 2015</td>
<td>CMS releases RFC with proposed changes to 2017 Star Ratings</td>
</tr>
<tr>
<td>February/April 2016</td>
<td>CMS finalizes 2017 Star Ratings in ANCL and Final Call Letter</td>
</tr>
<tr>
<td>October 2016</td>
<td>CMS releases ratings for 2017 on Plan Finder</td>
</tr>
<tr>
<td>2014</td>
<td>2015</td>
</tr>
<tr>
<td>May/June 2014</td>
<td>CMS proposes final measure set for 2017 in the proposed IPPS, with 60 days to comment</td>
</tr>
<tr>
<td>Late August 2014</td>
<td>CMS releases final IPPS rule and finalizes measures for 2017 ahead of 2015 data collection period</td>
</tr>
<tr>
<td>CY 2015</td>
<td>Hospitals report on quality measures/ CMS collects data (“Data Collection/Performance Period”)</td>
</tr>
<tr>
<td>October 2016</td>
<td>CMS releases ratings for 2017 on Hospital Compare</td>
</tr>
</tbody>
</table>

2017 MA Star Ratings Payment Implications FY2017

2017 Hospital IQR Payment Implications FY2017
Additionally, because measures are not finalized until after the performance period, plans may invest in quality improvement efforts targeted at a measure likely to be included in the MA Star Ratings (e.g., based on signals from CMS, or assumption of continued inclusion of an existing measure) only to have CMS decide to omit, or significantly change, the measure after the performance period. This is especially significant given the timing of when plans must go through their budget planning process for a year, which happens in June, ahead of when CMS may announce proposed measure set changes. Plans may miss an opportunity to budget for improvements ahead of a measures data collection period.

Although CMS does try to limit these types of “retroactive” changes (and even has guidelines for when it will remove a measure from the Star Ratings because of specification changes), these post-performance period changes create uncertainty for plans and hinder plans’ ability to make the most effective use of limited quality improvement resources. For example, in November of 2015, CMS proposed to exclude the High Risk Medication measure from the 2017 Star Ratings, though plans had already invested resources toward improving performance on the measure. Ultimately, CMS stepped back from this proposal because of significant stakeholder pushback about the amount of investment in strong performance on this measure. Nevertheless, this type of “retroactive” change could be proposed again because of the timing of finalizing measures.

Many Quality Incentive Programs Experience Significant Annual Changes in Measures and Methodology

CMS has introduced numerous annual changes to quality incentive programs, which not only require considerable resources from plans and providers to follow, monitor, and comply with each year, but also have the potential to destabilize year-to-year performance. These annual changes include things such as:

- Adding or retiring multiple measures in a given year;
- Adding or changing the weights of measures or measure categories;
- Removing or changing performance thresholds; and
- Changing the way improvement is measured and/or incorporated into the rating methodology.

As a result, these changes in methodology and measure sets can have meaningful implications for payments even absent no real changes in performance.

As an example, the Hospital VBP program has had relatively large changes to the number of measures included, the categories of measures included, and the weighting of categories of measures since it was first implemented (FY 2013). A recent Government Accountability Office (GAO) report examined trends in payment adjustments and quality performance under the Hospital VBP program. GAO found that while most hospitals had payment adjustments (both positive and negative) of less than 0.5 percent of applicable Medicare payments each year, the share of hospitals receiving adjustments of greater than 0.5 percent increased each year, growing from 7 percent in 2013 to 26 percent in 2015. Importantly, the GAO did not find real changes in the trends of quality performance due to the program. However, changes in the score calculations and weighting of measures may have changed the performance scores even when there was little change in actual quality.

MA plans have also experienced notable annual changes in the MA Star Ratings measures and methodology. For example, between 2011 and 2012, CMS retired 10 measures (nearly 20 percent of the total measure set in 2011) and added 10 new measures for 2012. Additionally, CMS introduced the measure weighting system that same year—prior to that the measures had not been weighted. While these kinds of significant, year-to-year changes present challenges on their own, the impact is exacerbated by the timeline employed by CMS; as discussed earlier, since CMS often announces changes after a performance period has occurred.

It is clear that while CMS continues to focus on evolving the quality incentive programs used in Medicare for the future, it will also continue to make annual changes to these programs. It is important that the MA Star Ratings program provide at least the same level of transparency and input (including formal comment opportunities) afforded to other quality incentive programs, and that all changes be implemented on a prospective, rather than retrospective, basis.
Measure Sets

The Number of Measures Included in the MA Star Ratings Could Be Reduced

The number and types of quality measures used in the different programs reviewed vary widely. Some of the programs used by CMS are more limited in scope and focus on improving quality in a specific area, such as readmissions or hospital-acquired conditions, and therefore have only one or a few measures. Other programs evaluate providers or plans over a wider array of quality metrics. The inclusion of too many measures in a quality incentive program may undermine quality improvement efforts by limiting the ability to target improvements, diluting the importance of individual measures, and makes it more challenging for stakeholders (including consumers) to understand drivers of quality.23

Penalty programs like HRRP, HAC and ESRD QIP tend to have fewer measures with 6, 6 and 11 measures each, respectively, in 2017. At the other end of the spectrum, reporting programs tend to include larger sets of measures, with over 271 measures for MIPS in 2017, and over 50 measures for IQR. However, it is important to differentiate these reporting programs, since any one provider does not need to report on all measures. Rather, providers select and report on a much smaller set of applicable measures. The number of measures included in Medicare VBP programs tends to be more moderate, with 14 measures included in the Hospital VBP in 2017 and for MIPS in 2017, physicians are required to report at least 6 measures, including at least one outcome measure.

ACOs, perhaps the model most similar to MA plans in terms of the overall responsibility for different aspects of patient care, report on 33 total measures. However, for scoring purposes these 33 measures are rolled up to 23 measures (6 of the CAHPS measures are combined into one measure and there are composite measures for diabetes and coronary artery disease).

The MA Star Ratings program includes the greatest number of measures for plans to report and be evaluated on among VBP programs. MA-PD plans are evaluated on 44 measures across Parts C and D. Additionally, as mentioned, CMS also collects and reports on plan performance on measures that are not part of the Star Ratings program, but may either be under consideration for inclusion in the measure set or may be retired from the measure set, on the Display Page. While not directly included in the MA Star Ratings, plans also have to focus attention on the large number of measures that are included on the Display Page.

While the volume of metrics means that no particular measure takes on too much importance and a wider variety of processes and outcomes can be included, having too many measures can dilute plans’ ability to target performance improvement to those areas that are most important and clinically meaningful.

Outcome Measures Remain Underrepresented in Many Measure Sets Including the MA Star Ratings

Across Medicare quality programs, CMS has consistently expressed an interest in focusing on outcome measures. Outcome measures are preferred by CMS because they measure changes in a beneficiary’s health as a result of the care that is provided to them, as compared to process measures which only evaluate whether certain activities take place.24 Outcomes measures are more likely to assess quality performance on measures of clinical relevance to Medicare beneficiaries.

CMS emphasizes outcome measures in quality incentive programs in one of two ways. The first is by increasing the number of outcome or intermediate outcome measures included in the measure sets. The second way, and perhaps the more relied upon method today, is through use of a weighting methodology that values outcome measures more than process or other types of measures. Part of the need to use weighting to increase the value of outcome measures, rather than simply adding more outcome measures, is that there is currently a limited number of outcome measures available for inclusion.

Despite an increasing focus on outcome measures or a move to make the measures more clinically relevant, process measures still tend to significantly outnumber the other types of measures in Medicare’s quality incentive programs. For example, the 2017 MIPS Quality measure set includes 271 measures total, of which 182 (67 percent) were process measures and only 73 (27 percent) were outcome or intermediate outcome measures (the remaining measures were efficiency measures, a patient engagement and experience measure, and structure measures). For the 2017 MA Star Ratings for MA-PDs, the number of process measures (16) was nearly double the number of intermediate outcome (6) and outcome (3) measures combined. However, because
of the weighting methodology CMS employs, the outcome and intermediate outcome measures account for approximately 34 percent of the Star Ratings, even though the number of outcome measures account for just 21 percent of the total measures.

The table below shows the number of measures by type in the MA Star Ratings as well as the weighted contribution to the overall score.

**Table 7. MA Measure Type as Percent of Total Measures, Unweighted and Weighted**

<table>
<thead>
<tr>
<th>MA Star Rating Measure Type</th>
<th>2017 Measure Count</th>
<th>Percent of Total Measures</th>
<th>Weighted Measure Value*</th>
<th>Weighted Measure Percent of Total Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>16</td>
<td>36%</td>
<td>16</td>
<td>20%</td>
</tr>
<tr>
<td>Access</td>
<td>7</td>
<td>16%</td>
<td>10.5</td>
<td>13%</td>
</tr>
<tr>
<td>Experience</td>
<td>10</td>
<td>23%</td>
<td>15</td>
<td>19%</td>
</tr>
<tr>
<td>Intermediate Outcome</td>
<td>6</td>
<td>14%</td>
<td>18</td>
<td>23%</td>
</tr>
<tr>
<td>Outcome</td>
<td>3</td>
<td>7%</td>
<td>9</td>
<td>11%</td>
</tr>
<tr>
<td>Improvement</td>
<td>2</td>
<td>5%</td>
<td>10</td>
<td>13%</td>
</tr>
</tbody>
</table>

*Note this does not reflect that new measures all receive a weight of 1 their first year no matter their type.

Outcomes measures are especially important in moving quality measurement toward clinically meaningful measures. Despite the efforts to include more outcome measures or to value them more heavily in score calculations, these measures—which assess actual improvements in beneficiary health status—are still underrepresented.

**Quality Incentive Programs Do Not Adequately Account for Differences Due to Socioeconomic Status**

A key issue for a number of quality incentive programs is whether the quality measurement system sufficiently addresses differences in performance due to plans or providers serving higher proportions or disproportionate numbers of beneficiaries who are lower income or have lower socioeconomic status (SES). Plans or providers serving higher proportions of beneficiaries with lower socioeconomic status may have distinct challenges in quality performance that may lower their performance comparatively. Under quality incentive programs, the potential to remove resources from plans and providers that disproportionately serve these populations and may already receive lower reimbursements may mean they have even less resources to reinvest in quality improvement.

Ideally, measures—or the ultimate ratings—impacted by socioeconomic factors would be adjusted to better account for the influence of SES on performance. However, many measures and rating systems do not yet account for SES and, as a result, likely disadvantage plans or providers that serve these populations. While NQF has recently revised their position on adjusting measures for SES, stating that some measures may need adjustments, this is a relatively new position and has not been widely adopted by CMS or measure developers yet.25

Further, the Department of Health and Human Services (HHS) Assistant Secretary for Planning and Evaluation (ASPE) recently published a report on the impact of low SES on Medicare’s VBP programs, including MA.26 This report clearly describes the significant impact of social risk factors on quality ratings and VBP across the Medicare program. Two key findings in this report were:

1. Beneficiaries with social risk factors had worse outcomes on many quality measures, regardless of the providers they saw, and dual enrollment status was the most powerful predictor of poor outcomes; and

2. Providers that disproportionately served beneficiaries with social risk factors tended to have worse performance on quality measures, even after accounting for their beneficiary mix.

Specific to MA, the ASPE report found that MA contracts with a high proportion of members with social risk factors generally did worse on overall quality scores, and were much less likely to receive quality bonus payments. The findings also suggested that the CAI did not sufficiently address the issue for plans and there needed to be a more significant and longer term adjustment.27
In the meantime, one possible approach CMS could take is to make the adjustment amounts under the CAI more meaningful to impacting overall Star Rating scores.

The lack of an SES adjustment for most measures included in the hospital quality incentive programs, particularly under the HRRP program, has drawn continued criticism and concern. The lack of an SES adjustment for most measures included in the hospital quality incentive programs, particularly under the HRRP program, has drawn continued criticism and concern. Numerous studies have demonstrated that hospital readmission rates are influenced by the SES characteristics of the population that the hospital serves. Studies have also demonstrated that safety net hospitals tend to have higher readmission rates. Additionally, there have been similar concerns raised about the Hospital VBP program. One recent review of the hospital pay-for-performance programs found that hospitals most likely to be safety net hospitals (those receiving the highest amount of disproportionate share payments) were two times more likely to receive a payment penalty under the VBP compared to hospitals less likely to be safety net hospitals.

Similarly, the need for adjustments to physician quality measures has also been raised. In recent NQF guidance on the measures under consideration for the MIPS program, one of the overarching themes was the need to better account for the impact of patients’ SES on quality performance. Specifically, they noted that “[t]he impact of patients’ SES and other demographic factors on measure results should continue to be explored, and it is important to take into account whether providers are caring for high-risk populations.”

The lack of an adjustment for SES for many of the MA and Part D Star Ratings measures has also been a serious concern for many MA plans. While some of the measures, like CAHPS measures, are case mix adjusted, other measures are not adjusted to account for the SES or disabled status of enrollees. There have been a number of studies demonstrating that plans that serve a higher proportion of low SES enrollees tend to have lower overall Star Ratings.

In response to concerns raised by plans, CMS conducted a study which showed a moderate within contract impact to performance for certain measures resulting from SES and disability status. CMS has reached out to measure developers to see if measure adjustment is warranted, but in the meantime, CMS continues to make the interim adjustment known as the CAI to the MA Star Ratings scores to account for differences in performance based on their research on within contract differences for certain measures. However, as noted above, the CAI has a very limited effect for most MA plans and most plans are concerned that it does not adequately address the issue.

**Scoring Methodology**

**Quality Incentive Programs’ Methodologies Are Often Overly Complicated**

In some of the quality incentive programs reviewed, the methodology for calculating the quality scores is quite complicated, raising questions about whether this complexity creates disincentives for participation (where voluntary) or confusion as to how the overall scores are computed. Further, the complexity in assigning scores can attenuate the connection between measure-level improvements and overall quality score. As a result, the burden associated with participation may not be commensurate with the payment incentives.

A number of factors contribute to the complexity of a scoring methodology, including the number of measures included, how performance scores are assigned for the measures, how the overall score is calculated including any weighting methodology, and how adjustments to scores (e.g., to account for low SES or to reward high performing plans or providers) can impact the final performance scores.

As shown in Figure 4, the MA Star Ratings has a particularly complicated methodology for several reasons:

- A large number of measures are included in the ratings, particularly for MA-PD plans.
- The cut points, or thresholds, for assigning Star Ratings to the individual measures are determined annually and after the data are collected, rather than prior to the measurement period.
- Measures receive varying weights in the final scoring, dependent on their type (e.g., process, outcome).
- The final scoring includes a complicated improvement measure, which factors in year-over-year performance on a subset of the measures but holds some plans harmless for consistently high performance.
• Additional adjustments are made before arriving at the final score including the CAI adjustment for differences in low-SES enrollment across contracts and application of a “reward factor” for plans with consistently high scores across multiple measures.

These adjustments add complexity and make it difficult for plans to relate improved performance on any one measure with improvements in their overall Star Ratings scores.

Figure 4. MA-PD Star Ratings Score Calculation Overview

Other programs reviewed had similar issues with complexity. Notably, the physician quality incentive programs (PQRS, VBPM, and Meaningful Use) have been considered extremely complex and onerous and have seen low participation rates by providers. For example, in 2015, 49 percent of physicians still did not participate in reporting on quality through PQRS and 40 percent received negative payment adjustments of 1.5 percent. Lack of participation is largely because of the complexity of participation but there are real questions about how much physicians can move quality performance at an individual level, especially under a system they may not fully understand. Some of the sweeping changes that were enacted under MACRA were intended, in part, to address the complexity of the current PQRS, meaningful use, and VBPM requirements.

More specifically, MACRA aims to streamline the requirements associated with physician quality incentive programs and to simplify the process to encourage greater participation and engagement. Congress enacted MACRA with two tracks for physician participation—one tied to quality performance under MIPS and the other tied to participation in alternative payment models. Despite the intention of MACRA to simplify current requirements, initial reactions from the provider community suggest that many physicians will find the new requirements as complicated as or more complicated than the current system.

Another issue with quality measurement complexity is the issue of investing in and developing systems to track performance or assist in identifying areas for performance improvement. While this may be an issue for providers generally, ACOs are also...
grappling with the complexity of their score development process and tracking measure performance. For instance, if ACOs do not, or cannot, invest in tracking measures that are claims-based, they may be missing opportunities to improve quality. These complexities may have disadvantaged some ACOs and even led to their disqualification from sharing in savings despite lowering costs.\(^4\)

Overall, the significant processes involved in assessing quality performance and determining quality scores may make it difficult for plans and providers to effectively monitor and impact performance. In part this is due to the attenuation of measure performance on overall scores, where the more complicated the process the more difficult for individual measure improvements to meaningful move a score. Plans and provider must also invest significant time and resources to understanding these quality evaluations.

Unlike Other Programs, Measuring Improvement in the MA Star Ratings Can Penalize Overall Performance Score

The MA Star Ratings include improvement measures that assess plans’ improvements on Part C and Part D measures. However, if plans perform poorly on the improvement measures, they can reduce a plan’s overall Star Rating—this is due in large part to the high weight assigned to the improvement measures in the overall Star Ratings.

This situation is unique among the quality incentive programs examined. Other programs reviewed compare measurements of performance achievement to performance improvement and take the higher score. Therefore, improvement is only incorporated into a score if it increases the performance score.

As an example, the Hospital VBP—used as the model for many of the VBP systems in terms of incorporating improvement (SNP VBP, HHA VBP)—compares measure performance to previous year’s performance as well as to other hospitals’ performance. In this way, the Hospital VBP scoring process evaluates both achievement and improvement for each measure and the hospital receives the higher of the two scores—achievement or improvement. As a result, the program evaluates improvement on a measure-by-measure basis and rewards providers that have significant improvements, but it does not necessarily penalize providers that did not improve or even declined compared to their baseline performance, if that performance is still favorable relative to other hospitals’.

The improvement measures for the MA Star Ratings work differently. The improvement measures for both Part C and Part D compare performance over two years on a number of measures. These comparisons result in the assignment of a Star Rating for each of the Part C and Part D improvement measures. These measures are worth 5 times the process measures and are therefore the most heavily weighted measures in the MA Star Ratings. Therefore, a low rating on the improvement measure(s) could impact up to 13 percent of the plans’ overall Star Rating (see Table 6 above). CMS does apply a “hold harmless” approach for plans that see a reduction in measure scores but still have a 5-Star Rating overall.

Use of Pre-Set Measure Thresholds Uncommon Among Quality Incentive Programs but May Have Value

While most of the quality incentive programs reviewed do not include pre-set thresholds for performance, many, like the Hospital VBP program, measure improvement by comparing the performance period against a baseline period. In this way, while not necessarily a benchmark of all other facilities’ performance, providers can benchmark against previous performance. And significant improvement from year-to-year can be used as the measure score, if more favorable than performance relative to other facilities’.

Under MACRA, CMS will also provide thresholds for the total composite score. These performance thresholds will be determined annually as the mean or median of the MIPS scores for all eligible providers in a prior period as selected by CMS. Initially, in the first years of the MIPS program, CMS will not have the historical data to develop the performance thresholds but will instead use a combination of performance level on other quality incentive programs including meaningful use.

For MA, CMS used to announce pre-determined four Star performance thresholds for certain measures. These thresholds helped plans to target higher quality performance on these measures. Many plans also used these targets to help incentivize provider performance. CMS stopped using the thresholds beginning in 2016. Now, CMS sets Star Ratings levels primarily based on cutpoints established annually by statistical analyses such as “clustering” of plans’ performance. In this way, plans are graded on a curve each year for most measures.
While this may encourage continual improvement efforts from plans, it makes it difficult for plans to target performance. Further, it does not guarantee that an improvement over a previous year’s performance will lead to an improved score since overall plan performance may shift. It also may make it difficult to work with providers to better drive improvements in performance.

**Consumer Transparency**

While quality measurement is used to link payments to value and thus incentivize quality improvements, it is also used to encourage beneficiaries to select higher value care. To do so, Medicare provides beneficiaries with information about the quality and value of the care they select and receive. Most commonly, CMS provides consumers information on quality via one of their “Compare” websites, such as Hospital Compare, Plan Compare, or Physician Compare. Although these sites are meant to provide consumers much needed quality and sometimes cost information, many of them are also considered to be poorly organized, not entirely consumer friendly, and lacking consistently reliable data. As a result, CMS continues to try to improve them.

How quality information is presented to consumers on these websites varies. Some programs, like the MA Star Ratings, Dialysis Facility Compare and the Home Health Agency ratings, are shown using a five-Star system which combines multiple performance measures into an overall score ranging from one to five Stars. Other of CMS’ websites, particularly Physician Compare, presents a wide array of specific quality metrics but lack an easily interpretable overall rating score. Further, these websites may often provide a variety of quality information to consumers, such as Hospital Compare.

Hospital Compare includes information for over 4,000 hospitals on more than 100 quality measures (though not all hospitals have information for all measures). These include an individual hospital’s scores on HRRP, HAC, and the VBP programs as well as IQR and OQR reported measures. The website also allows consumers to compare up to three hospitals at a time across the measures. CMS recently announced that it will combine 62 of the existing measures available on Hospital Compare to create a “unified” five-Star rating.

On Physician Compare, CMS presents limited quality information, but does indicate whether physicians participate in various incentive programs such as PQRS, VPM, and EHR reporting. CMS also has made some information on quality performance for some group practices available but this is very limited. Additionally, ACO quality performance information, including CAHPS data, is available on the Physician Compare website.

In contrast, the MA Star Ratings have long been available to consumers and the use of the ratings for consumer information pre-dated their use for payment. CMS established the quality rating system in 2006 to help beneficiaries in their selection process for plans. Plans’ quality information is summarized in overall and measure-by-measure 5-Star ratings. CMS also uses special indicators to distinguish plans with the highest quality or that are low performing plans. One shortcoming of Plan Finder is that it presents information about FFS coverage and costs to compare with MA plans, but it does not provide comparable information on FFS quality.
RECOMMENDATIONS

Based on the findings presented in this comparison of Medicare quality incentive programs, several opportunities to strengthen the current MA Star Ratings program emerge. Below we briefly discuss some key recommendations for each of the four main areas we compared.

Program Processes

1. CMS should implement and use a formal rulemaking process for the MA Star Ratings program that provides a 60-day comment opportunity, has earlier and more thorough vetting of new measures, and finalizes measure sets and scoring methodology in advance of the data collection and performance periods.

One of the most critical findings from this evaluation is that the MA Star Ratings program changes are not announced or implemented in the same way as for other Medicare quality incentive programs. This is particularly problematic in that plans do not know the full set of measures that will be included in their performance scores ahead of the performance and measurement period. Further, CMS can make changes after the performance periods which impact a plan’s score. CMS should apply all modifications on a prospective basis and finalize measures and their methodologies prior to the start of the measurement period. The MA Star Ratings is the only program where the measures are not announced in advance of their measurement.

Measure Sets

2. CMS should reduce the number of measures included in the MA Star Ratings, while focusing on clinically meaningful outcome measures to drive greater improvements in health outcomes.

MA plans are evaluated on more measures than any other program, including ACOs. CMS should reduce the number of measures included in the MA Star Ratings to help plans better target their performance improvement efforts and resources. This change would also better align the MA Star Ratings with other Medicare value-based purchasing programs.

Also, CMS should focus on including clinically meaningful measures, especially outcome measures with a more direct link to the actions and outcomes that improve beneficiary care. In particular, we suggest that CMS attribute greater weight to data-driven outcome measures with objective clinical relevance than to measures of health status that are constructed from enrollee surveys.

3. CMS should continue to examine and test more meaningful adjustments to plans’ ratings to account for differences in the socioeconomic status of enrollees.

Despite the recent inclusion of an adjustment for SES through the introduction of the CAI for MA plans, more needs to be done to address performance differences due to the socioeconomic differences of plans’ members. This is especially important for plans serving high proportions of beneficiaries with low SES. The CAI did not have a material impact on the overall rating for the vast majority of plans; the outcome does not warrant the complexity of the process to make the adjustment.
Given the immaterial impact of the CAI, CMS should continue actively working with the measure developers, ASPE, plans, and other stakeholders to develop and implement a long-term solution that is sufficient and meaningful. If CMS truly intends to make an impact through an adjustment aimed at addressing the impact of low-SES, it should be on par with the increments of the reward factor such that the adjustment is more meaningful and could shift more overall scores.

**Scoring Methodology**

4. Given the complexity and multitude of quality incentive programs that providers face, CMS should continue to work towards greater alignment of measures and objectives between FFS and MA plans.

In order to make a meaningful shift in quality for beneficiaries, there should be alignment across the FFS and MA programs to promote consistency and to better focus improvement efforts. Most providers see both FFS and MA members and are becoming increasingly accountable for performance over an array of metrics. In order to lessen the burden on providers as well as to reinforce areas for needed improvements, CMS should use the same or similar measures and data sources when possible. This would allow greater emphasis on overall clinical goals and would not create competing or even conflicting quality indicators for providers.

5. CMS should review the methodology used for the improvement measures in Part C and Part D to ensure it rewards plans for improvements in performance, protects plans that are higher performing from being penalized, and promotes improvement opportunities for all plans including lower performing plans.

Consistent with other quality incentive programs in Medicare, the calculation and the application of the improvement measures for MA plans should be improved to ensure that they do not negatively impact higher performing plans and do benefit all plans that improve. In order to protect higher performing plans, CMS should re-evaluate its “hold harmless” provision so that plans are not penalized for performance at the 4 or 5 Star level. Finally, CMS should allow plans with 2.5 Stars to benefit from the improvement measures. Limiting the measure to only plans with more than 2.5 Stars goes against the objective of the improvement measure in encouraging and rewarding improvements in performance, particularly among lower-rated plans.

6. CMS should reinstate four-star thresholds for selected measures to more clearly identify desired levels of performance based on prior year trends, as well as expected and realistic improvement standards, and to encourage stable improvement objectives.

CMS has previously used four star thresholds in the MA Star Ratings to help plans target their performance improvement efforts. This not only provided plans with a stable target—as opposed to one that could shift from year-to-year due to differences in performances of all plans—but also allows plans to work with providers to achieve meaningful improvements in quality. Additionally, CMS has recognized the value of performance thresholds to help drive improvement in other quality incentive programs examined.
**CONCLUSION**

The MA Star Ratings program is a successful example of effectively encouraging quality improvement though payment incentives. Despite the relative success, there continue to be aspects of the program that could be improved. By examining the MA Star Ratings in the context of other Medicare quality incentive programs, we identified specific areas that could be strengthened to better incentivize plan improvements.

In comparing quality incentive programs across Medicare along key features, it is clear that the MA Star Ratings program is unique in the process by which program changes are announced and made. This is a critical issue for plans—in order to drive improvement they need greater transparency, clarity, and timeliness about program and measurement changes. Providing a more formal process that includes finalizing measures and methodology before performance and measurement periods would go a long way to strengthening the program.

Further, efforts to refine the measure set and methodology of calculating the MA Star Ratings could have meaningful impacts on incentivizing performance improvements. Specifically, focusing on generally reducing the number of measures while at the same time including more outcome measures that plans have a better ability to influence, providing thresholds for targeted improvements, and making more meaningful adjustments to account for plans serving higher proportions of SES are critical to continued quality improvement. It is difficult to make meaningful improvements when there are significant methodological changes each year or significant changes to the numbers and types of measures included in the programs.

Quality improvement can best happen when there are clear expectations. The recommendations presented suggest ways to enhance the MA Star Ratings so that it is more aligned with other Medicare quality incentive programs and continues to encourage plan quality improvements.

To aid Medicare beneficiaries in making a fully informed enrollment choice that is based on both cost and quality considerations, CMS should provide comparable measures of FFS quality for beneficiaries to consider alongside MA plans’ Star Ratings.

A key reason for evaluating and publicly releasing quality performance information is to help inform beneficiaries’ choice of plans and/or providers and to drive care toward those that are higher quality. Currently, beneficiaries are not able to make fully informed decisions with regard to quality because there is not sufficient information provided about FFS quality compared to MA plan quality. CMS should provide comparable quality information for FFS on Plan Compare so beneficiaries are better able to make informed decisions about their Medicare coverage.
APPENDIX A:
OVERVIEW OF PAY-FOR-PERFORMANCE PROGRAMS EXAMINED

<table>
<thead>
<tr>
<th>Program</th>
<th>Program Description</th>
<th>Covered Entities</th>
<th>Type of Incentive Program</th>
<th>Payment Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MA Star Ratings</strong></td>
<td>The MA/Part D Star Rating program assesses the quality of plans participating in the Medicare program. Plan quality is assessed over a variety of process, access and experience, and outcome (or intermediate outcome) measures. There are separate ratings for MA and Part D, but MA plans with Part D coverage receive a score for the combined set of measures. The quality ratings are only used to adjust payments to MA plans, not Part D plans.</td>
<td>MA and Part D Plans</td>
<td>Bonus payment for higher performance</td>
<td>5% county payment rate (10% in qualifying double bonus counties) and rebate percentages tied to Star Ratings</td>
</tr>
<tr>
<td><strong>Hospital Inpatient Quality Reporting (IQR)</strong></td>
<td>The Hospital Inpatient Quality Reporting (Hospital IQR) program penalizes hospitals that do not report designated quality measures by reducing the annual update to their payment rates.</td>
<td>IPPS Hospitals</td>
<td>Pay for reporting</td>
<td>One-fourth reduction of annual payment update for failure to report measures</td>
</tr>
<tr>
<td><strong>Hospital VBP</strong></td>
<td>The Hospital VBP program is a pay-for-performance program for inpatient hospitals that receive Medicare prospective payments. Under Hospital VBP, Medicare adjusts payments to hospitals based on either: • How well they perform on each measure compared to all hospitals, or • How much they improve their own performance on each measure compared to their performance during a prior baseline period.</td>
<td>IPPS Hospitals</td>
<td>Pay for performance</td>
<td>In 2017 2% withhold of base payments; funds are then redistributed to hospitals based on performance</td>
</tr>
<tr>
<td><strong>Hospital HAC</strong></td>
<td>Reduces payments to IPPS hospitals that rank in the worst performing quartile of hospitals with respect to hospital-acquired conditions (HACs).</td>
<td>IPPS Hospitals</td>
<td>Penalty for poor performance</td>
<td>1% of total IPPS payments including add-ons</td>
</tr>
<tr>
<td><strong>Hospital HRRP</strong></td>
<td>Reduces payments to acute care hospitals with excess readmissions that are paid under IPPS. The program initially focused on patients who were readmitted for selected high-cost or high-volume conditions, namely, heart attack, heart failure, and pneumonia, but now includes readmissions for COPD, elective total hip/total knee arthroplasty and coronary artery bypass grafting.</td>
<td>IPPS Hospitals</td>
<td>Penalty for poor performance</td>
<td>Up to 3% of operating payments</td>
</tr>
<tr>
<td><strong>Hospital Outpatient Quality Reporting</strong></td>
<td>The Hospital Outpatient Quality Reporting Program (Hospital OQR) is a pay for reporting program for outpatient hospital services. Requires outpatient hospitals to submit data on measures of the quality of care furnished by hospitals in outpatient settings or face reduction in OPPS payments.</td>
<td>OPPS Hospitals</td>
<td>Pay for reporting</td>
<td>2% reduction in OPPS payment update</td>
</tr>
</tbody>
</table>
## Opportunities to Strengthen the Medicare Advantage Star Ratings Program: A Comparison to Other Medicare Quality Incentive Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Program Description</th>
<th>Covered Entities</th>
<th>Type of Incentive Program</th>
<th>Payment Implications</th>
<th>Quality Information Available to Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PQRS</strong></td>
<td>The Physician Quality Reporting System (PQRS) applies a negative payment adjustment to individual eligible professionals and PQRS group practices who do not satisfactorily report data on quality measures for Medicare Part B Physician Fee Schedule (MPFS) covered professional services. Note: Physicians or practices that do not report also are penalized under the physician VBP program.</td>
<td>Physicians and physician practices</td>
<td>Pay for reporting</td>
<td>Total of 4%: 2% penalty for PQRS failure to report and 2% penalty under VBP</td>
<td>Yes-Physician Compare</td>
</tr>
<tr>
<td><strong>Physician VBPM</strong></td>
<td>The Value Modifier provides for differential payment to a physician or group of physicians under the Medicare Physician Fee Schedule based on the quality of care furnished compared to the cost of care during a performance period. The Value Modifier is used to adjust Medicare payments to non-physician eligible professionals, in addition to physicians.</td>
<td>Physicians and physician practices</td>
<td>Pay for performance</td>
<td>In 2017: For groups of 10-99 practitioners: adjustment of -2% to +2%. For groups of 100 or more: adjustment of -4% to +4%.</td>
<td>Yes-Physician Compare</td>
</tr>
<tr>
<td><strong>Meaningful Use</strong></td>
<td>Meaningful Use is a payment adjustment for eligible providers based on their use of electronic health records (EHR).</td>
<td>Physicians and physician practices</td>
<td>Penalty (incentive payments were used until 2015)</td>
<td>3½% in 2017 (shift to MIPS in 2018)</td>
<td>Yes-Physician Compare</td>
</tr>
<tr>
<td><strong>MIPS</strong></td>
<td>Merit-Based Incentive Payment System (MIPS) will replace PQRS, physician VBPM, and meaningful use beginning in 2019. Physicians will receive positive or negative adjustments depending on performance across a range of categories including: Quality, Resource Use, Advancing Care Information, and Clinical Practice Improvement Activities.</td>
<td>Physicians and physician practices</td>
<td>Pay for performance</td>
<td>Will impact 4% of payments in 2019 and up to 9% by 2022 (high-performing providers can receive higher payment updates)</td>
<td>TBD</td>
</tr>
<tr>
<td><strong>ESRD QIP</strong></td>
<td>The ESRD QIP reduces payments to ESRD facilities that do not meet certain performance standards.</td>
<td>Dialysis facilities</td>
<td>Penalty</td>
<td>Up to 2% penalty</td>
<td>Yes-Dialysis Facilities Compare</td>
</tr>
<tr>
<td><strong>HHA Quality Reporting</strong></td>
<td>Home health agencies (HHAs) are required to submit quality data, and payments are reduced for agencies that do not submit the information.</td>
<td>Home Health Agencies</td>
<td>Pay for reporting</td>
<td>2% penalty for failure to report</td>
<td>Yes-Home Health Compare</td>
</tr>
<tr>
<td><strong>HHA VBP Pilot Program</strong></td>
<td>Pilot program for home health agencies in 9 states that ties Medicare payments to quality performance. The pilot runs from January 1, 2016 through December 31, 2022. 2016 is the first performance year; payment implications begin in 2018.</td>
<td>Home Health Agencies (in 9 states)</td>
<td>Pay for performance</td>
<td>Bonus or penalty payments of 5-8% beginning in 2018 depending on performance</td>
<td>NA</td>
</tr>
<tr>
<td><strong>ACO Quality Measure Program</strong></td>
<td>In order to receive any shared savings amounts, ACOs must also meet certain quality performance standards.</td>
<td>ACOs</td>
<td>Shared Savings Eligibility</td>
<td>Determines shared savings availability</td>
<td>Yes-Physician Compare</td>
</tr>
<tr>
<td><strong>SNF VBP</strong> (begins in FY2019)</td>
<td>Skilled Nursing Facilities will have payment tied to performance based on either achievement or improvement.</td>
<td>Skilled Nursing Facilities</td>
<td>Penalty</td>
<td>Up to 2% of Part A payments</td>
<td>Yes-To be posted on Nursing Home Compare</td>
</tr>
</tbody>
</table>

Opportunities to Strengthen the Medicare Advantage Star Ratings Program:
A Comparison to Other Medicare Quality Incentive Programs

---

**Anthem Public Policy Institute**
END NOTES


7. Ibid.


9. While still a sub-regulatory process, Congress recently passed legislation changing the timing of the release of the Advance Notice and lengthening the comment period. In December of 2015 Congress passed legislation that would establish, beginning in 2017, a revised annual schedule for the development of Medicare Advantage (MA) payment rates. Beginning with the 2018 Advance Notice (released in 2017), the Advance Notice will be released 60 days ahead of the Final Rate Announcement and stakeholders will have a 30 day comment period. Prior to this change, the Advance Notice was released 45 days ahead of the Final Rate Announcement with a two week comment period for stakeholders. The Draft Call Letter has been released with the Advance Notice the last few years; if CMS continues to releases them together then the Draft Call Letter has the same timeline and comment opportunity.

10. See endnote 7.


Opportunities to Strengthen the Medicare Advantage Star Ratings Program: 
A Comparison to Other Medicare Quality Incentive Programs


27. Ibid.


About the Anthem Public Policy Institute
The Anthem Public Policy Institute was established to share data and insights to inform public policy and shape the health care programs of the future. The Public Policy Institute strives to be an objective and credible contributor to health care innovation and transformation through publication of policy-relevant data analysis, timely research, and insights from Anthem’s innovative programs.

About Anthem, Inc.
Anthem is working to transform health care with trusted and caring solutions. Our health plan companies deliver quality products and services that give their members access to the care they need. With over 73 million people served by its affiliated companies, including more than 40 million enrolled in its family of health plans, Anthem is one of the nation’s leading health benefits companies. For more information about Anthem’s family of companies, please visit www.antheminc.com/companies.