AMA
AMERICAN MEDICAL ASSOCIATION

JAMES L. MADARA, MD
EXECUTIVE VICE PRESIDENT, CEO

ama-assn.org
t (312) 464-5000

January 2, 2018

The Honorable Seema Verma
Administrator
Centers for Medicare & Medicaid Services
U.S. Department of Health and Human Services
Hubert H. Humphrey Building, Room 445–G
200 Independence Avenue, SW
Washington, DC 20201

Re: CY 2018 Updates to the Quality Payment Program; and Quality Payment Program: Extreme and Uncontrollable Circumstance Policy for the Transition Year (CMS-5522-FC and IFC)

Dear Administrator Verma:

On behalf of the physician and medical student members of the American Medical Association (AMA), I am pleased to offer our comments to the Centers for Medicare & Medicaid Services (CMS) on the 2018 Quality Payment Program (QPP) final rule and interim final rule. The AMA supports many of the policies finalized by CMS for 2018, and appreciates that the agency is working to create a new program that reduces burden, while also recognizing the need for flexibility given the variety of clinicians participating in the QPP. However, we strongly oppose CMS' decision not to finalize the cost performance category weight at zero percent in 2018. Given that new episode-cost measures are still being developed, physicians will be evaluated on two flawed cost measures carried over from the Value Modifier (VM) program in 2018, which will hold many physicians accountable for costs that they had no control over while leaving many others with no applicable cost measures.

The AMA also offers comments on topics CMS is specifically seeking feedback on in this final rule, including the low-volume threshold, bonus points, the Advanced Payment Model (APM) Certified Electronic Health Record Technology (CEHRT) requirement, and the unforeseen and uncontrollable circumstances hardship exemption. Also included are the AMA's analysis and recommendations on improving benchmarking methodology for measures in the Quality performance category. For the AMA's comments on other issues, please see our 2018 QPP proposed rule comment letter. We are committed to working with CMS to continue to improve the QPP.

I. Policies the AMA Supports

The AMA commends CMS on finalizing a number of policies recommended by the AMA for the 2018 performance year. Specifically, we strongly support the following changes for the 2018 QPP performance year:

- The expansion of the low-volume threshold to individuals and groups that have Medicare Part B allowed charges less than or equal to $90,000 or that provide care for 200 or fewer Part B-enrolled Medicare beneficiaries. However, we also continue to urge CMS to allow physicians to opt-in to the Merit-Based Incentive Payment System (MIPS) program if they wish to participate in future years.

- The ability for small groups and solo practitioners to form virtual groups. We also believe that physicians should have maximum flexibility in the formation of virtual groups and will continue to work with CMS to encourage the development and success of virtual groups.

- New bonus points for small practices and physicians who treat complex patients.

- CMS' policies within the Advancing Care Information (ACI) performance category to extend CEHRT flexibility for performance year 2018 and provide a hardship exemption for small practices.

- Several policies CMS finalized in the Quality performance category for the 2018 performance year including not increasing the number of measures physicians must report, and eliminating cross-cutting measures from many of the specialty measure sets.

- CMS' decision to maintain the reporting and performance requirements within the Improvement Activities category to provide stability within the MIPS program.

- CMS' decision to extend the 8 percent revenue-based nominal amount standard for APMs for an additional two years, and allow Other Payer APMs to use the revenue-based standard.

- CMS' plans to develop a demonstration project to test the effects of allowing credit for participation in Medicare Advantage APMs starting in 2018.

- The gradual phase in of the lower financial risk standard for medical homes which ranges from 2.5 percent to 5 percent of revenues.

II. Cost Category Concerns

While there are many aspects of the final rule that we support, the AMA is extremely disappointed that CMS chose to reverse its earlier decision to assign a zero weight to the cost category again in 2018. As we have repeatedly observed, the two measures that physicians will be judged on in 2018 are highly flawed, often irrelevant, and jeopardize access to care for patients with high cost conditions. Perversely, under these measures, physicians can even be punished for costs that were incurred before they met the patient.

In reversing the 2018 proposed rule proposition to set the 2018 performance year weight at zero rather than the previously finalized 10 percent, CMS said it was acting out of continuing "concerns about the level of familiarity and understanding of cost measures among physicians." The agency further noted that taking an extra year to prepare for implementation of the cost category would provide it with time to educate physicians about cost measurement and to develop new measures that are based on episodes of

care and have involved significant input from physicians and other clinicians. CMS was worried, however, that this would create a "sharp increase" to the Medicare Access and CHIP Reauthorization Act (MACRA) required weight of 30 percent in 2021 and sought comments on the relative merits of a zero versus a 10 percent cost category weight.

The AMA believes that there is still a need for more time and more outreach to educate physicians about how cost measures work and how likely they are to affect a particular physician's total score. Currently, however, the only way physician practices can learn this is by acquiring an Enterprise Identity Management (EIDM) account, then pulling and reviewing a Quality and Resource Use Report (QRUR), and then trying to calculate how this would affect their overall MIPS score. Acquiring a QRUR is a tedious and often frustrating task that should not be relied upon as the only means of familiarizing physicians with either the current cost measures or those that are under development for future use. It also may not be clear to physicians how these measures will affect their overall score.

We are aware of and applaud CMS' efforts to develop a new tool that physicians can use to track their performance under various MIPS categories. However, access to the tool will still require an EIDM account and it is not yet clear that the tool will include up-to-date cost measure calculations. Without this data, physicians will remain in the dark about how cost measures will impact their score and many may be unpleasantly surprised to discover that they are no longer eligible for a bonus and may even be facing a penalty.

This, in and of itself, would have been a good reason for keeping the cost category weight at zero in 2018. However, the most important reason for delaying implementation of the cost category is the current lack of any reliable and valid cost measures. For reasons that were laid out in more detail in our comments on the proposed rule, the AMA considers the 0.4 percent reliability standard CMS has set for its cost measures to be unacceptable, and is opposed to the two VM carry-over measures CMS intends to use for the second year of MIPS Total Cost of Care (TCC) and Medicare Spending Per Beneficiary MSPB.

We are pleased that a CMS contractor, Acumen LLC is working on improvements in the two VM measures and developing new episode-based measures with the assistance of clinical panels. However, none of these improvements could be implemented before 2019 at the earliest. **Consequently, any feedback based on the current measures will be of little or no value because they do not address costs that physicians can actually control, and because the measures are likely to change in the next several years.**

Even more importantly, without the planned improvements, the TCC and MSPB measures will compare physicians and influence the distribution of Medicare penalties and rewards based on a reliability standard that is rejected by most experts and on cost measures that have consistently discriminated against physicians treating Medicare's frailest patients. CMS stated it decided to weight the cost category at 10 percent in 2018 so that physicians would have a warning signal before the weight jumped to the statutorily mandated 30 percent in 2019. To benefit from this warning, however, physicians will need to reduce or constrain resource use and because they do not control all the costs attributed to them under the TCC and MSPB measures, the easiest way to avoid a penalty will be to avoid high cost patients. While the complex bonus may mitigate this incentive somewhat, the bonus will be temporary and ideally the patients will not.

The AMA believes that getting the measures right would have been better than applying bad measures a year earlier than necessary just to provide physicians a warning of what is to come. **To punish or reward physicians based on the currently unreliable measures in the interim may indeed call attention to the measures, but it risks physician alienation and threatens continued access to adequate medical care for some Medicare beneficiaries.**

III. General MIPS Provisions

Low-Volume Threshold

The AMA strongly supports the expansion of the low-volume threshold to individuals and groups that have Medicare Part B allowed charges less than or equal to $90,000 or that provide care for 200 or fewer Part B-enrolled Medicare beneficiaries. The AMA had strongly advocated for additional relief for small and rural physicians, and we applaud CMS for finalizing this policy which will reduce the impact of MIPS on physicians in small and rural practices.

While we support the expansion of the low-volume threshold, as we have stated in previous comment letters, we believe that physicians should be allowed to opt-in to the MIPS program if they wish to participate in future years. This will allow those physicians who are ready to report or wish to gain experience with the program to learn the MIPS requirements and have an opportunity to earn an incentive payment. We understand the complexity involved with offering this option to clinicians; however, we urge CMS to develop this option in the future. We also encourage CMS not to limit optional participation to only those physicians who meet or exceed one, but not all of the low-volume threshold determinations. Rather, CMS should allow all of those below the low-volume threshold the ability to opt in and participate in MIPS in future years.

Multiple Submission Mechanisms

In the 2018 QPP proposed rule, CMS proposed to allow physicians to submit data on measures and activities through multiple submission mechanisms for a single performance category. Although the AMA supported CMS' efforts to provide physicians additional flexibility in MIPS reporting, we were concerned that as CMS proposed, scoring measures across multiple submission mechanisms might make reporting quality measures more complex, costly, and burdensome for physicians.

Specifically, the AMA was concerned that physicians who had fewer than six measures available under one submission mechanism may be required to use a second submission mechanism in order to receive a maximum quality score. CMS clarified in the final rule that although it finalized allowing multiple submission mechanisms for a single performance category beginning in 2019, it will not require that physicians submit via multiple submission mechanisms if a physician has fewer than six measures to report through a single submission mechanism. **The AMA appreciates this clarification, and agrees it will eliminate some of the unintentional burden this policy might have created for physicians.**

IV. MIPS Scoring Methodology

Throughout the final rule, CMS seeks comments on how to simplify the scoring system while recognizing differences in clinician practices. The AMA appreciates CMS' recognition that the overall scoring methodology for the MIPS program should be simplified and that flexibility is needed to allow for the wide variety of physician specialties and practice sizes. As we have commented previously, if physicians do not comprehend how CMS is calculating MIPS scores, they are likely to view the program as unfair

and may be subject to financial penalties solely due to confusion rather than their actual performance. While we understand part of the scoring methodology complexity is due to statutory language and the requirement for a composite score, we believe there are several changes that could be made to the scoring methodology to simplify the program, including those highlighted below which CMS sought feedback on in this final rule.

Bonus Points

The AMA supports the new bonus point opportunities finalized by CMS, including additional points for small practices and providers who see complex patients. The AMA appreciates CMS' efforts to recognize and reward physicians who face unique challenges in the MIPS program. However, we urge CMS to make the bonus points permanent.

In the final rule, CMS states that it intends the bonus points to serve as a short-term solution, and may not provide the bonus points in future performance periods. Temporary bonus points not only create complexity, but will artificially inflate the performance threshold for participants. These participants will then be disadvantaged in future program years when the bonus points are removed or reduced, essentially creating a greater hardship for the categories of participants who need the most assistance. Therefore, in addition to encouraging CMS to make the bonus points permanent, we also believe that the bonus points should not be factored into setting future performance thresholds.

Score Outcome Measures and High Priority Measures Equally

Under the current scoring rules, if a physician reports on additional outcome measures they receive two achievement points, but if a physician reports on additional high-priority measures they only earn one achievement point. The inconsistency between the scoring rules is confusing and CMS does not clearly distinguish the difference on the QPP website, as both are designated as high priority measures.

In addition, to fully satisfy the quality requirements, a physician must report on an outcome measure. If they do not believe there is an applicable outcome measure for their practice, and they pick a high priority measure as an alternative, but CMS believes an outcome measure is in fact applicable to their practice, they are penalized in their scoring. **CMS should simplify scoring in the Quality performance category by making outcome measures and high priority measures optional, and awarding the same number of bonus points for reporting outcome and high priority measures.** Regardless of whether CMS maintains the outcome measure requirement, outcome measures and high priority measures should be scored the same.

Facility-Based Measurement

While the AMA supports allowing physicians to select a facility-based measurement option in 2019, we continue to urge CMS to reduce the facility-based measurement point floor. Given the 30 percent floor that CMS finalized for physicians choosing to report via the facility-based measurement option, any physician who selects the facility-based measurement option would automatically score above the performance threshold regardless of the performance of his or her facility. The AMA believes this high point floor is unfair to non-facility based clinicians and urges CMS to reduce the point floor for physicians opting into facility-based measurement.

Physicians opting into facility-based measurement already have some advantages in the Quality performance category, such as the requirement that if a physician elects facility-based measurement, but also submits quality data through another mechanism, CMS will use the higher of the two scores for the

Quality category. We are also concerned that if the point floor is maintained, facilities may not be incentivized to invest additional resources into physician-level quality reporting tools, which would create problems for physicians that choose to report separately from the facility. Specifically, there would be no incentive for the facility to coordinate with individual physicians or specialties on meaningful quality measures when the physician can achieve a score higher than the performance threshold regardless of their performance in the Value Based Purchasing (VBP) program. Therefore, CMS should reduce the 30 percent floor in the Quality performance category for physicians electing to use facility-based measurement in 2019, to ensure the program is equitable for both facility and non-facility based physicians.

Benchmarking of Quality Measures

The AMA is extremely concerned with the lack of transparency in the methodology used for creating quality measure benchmarks. It appears benchmarks for the 2017 performance period were created using data from a small number of physicians, and it is not clear whether the scores truly reflect performance. The benchmarks were also developed based on 2015 Physician Quality Reporting System (PQRS) reporting data, and if CMS follows the same timeline in 2018, MIPS benchmarks will be based off of 2016 PQRS reporting data. These data are problematic, as they were used in a different quality reporting program, under which physician were scored differently.

In addition, increasing the low volume threshold, which the AMA strongly supports, could also have an impact on MIPS benchmarks because a greater number of physicians will be exempt from MIPS, but might be included in the benchmarks using previous PQRS data.

In an attempt to examine and explore the current MIPS evaluation methodology, we compared physician performance under the current MIPS methodology to performance as determined by the Physician Compare five-star rating Achievable Benchmark of Care (ABC) methodology. The two methodologies resulted in inconsistent ratings and comparisons (see Appendix). In several instances, physicians deemed to be of similar quality by one methodology were classified as having different levels of quality by the other methodology. Our analysis does not attempt to determine which methodology more accurately reflects true differences in quality, but implies that at least one of the methodologies is lacking, and suggests further thought and testing is necessary. In addition, the differences between the two methodologies add to the complexity of the QPP program. In an attempt to improve and simplify the MIPS program, the AMA continues to explore and perform analyses on alternative MIPS benchmarking methodologies to help implement a more stable, evidence-based program. We welcome the opportunity to discuss the initial analysis provided in the Appendix further with CMS.

Furthermore, the following are areas that need additional exploration:

- *Quartile vs. Decile Spread vs. Star Ratings:* Reducing the spread from deciles should assist with strengthening the underlying data that form the quality measure benchmarks. CMS' methodology for creating MIPS benchmarks and calculating achievement points conflicts with the ABC methodology used to determine Physician Compare star ratings. Differences in the methodologies include how the top decile or benchmark is determined and how performance is distributed across the deciles. Based on our preliminary analysis, the ABC methodology could be a better approach than using deciles for setting MIPS quality benchmarks. This methodology may also provide more accurate data on "topped out" measures. The deciles may yield lower ratings and numbers than the five-star method due to the decile cap on "topped out" measures.

Therefore, it would be helpful if CMS could provide more information on the ABC methodology and analyze whether the Physician Compare methodology should be used to calculate MIPS benchmarks, as well as for Physician Compare star ratings. This approach would also create some consistency between how a physician scores points and achieves stars.

- *Sample Size:* **We are concerned that for some quality measures there may not yet be sufficient data to ensure that the results are reliable and reflective of true performance due to sample size.** We also question whether a minimum of 20 patients is a sufficient minimum sample for all measures, or whether CMS should determine on a measure by measure basis what is the appropriate sample size to set benchmarks.

- *Transparency:* **For greater transparency, it would be useful if the minimum, average and maximum number of eligible clinicians who successfully reported on a measure, and the number of patients included in the denominators, as well as the number of eligible clinicians who attempted but did not successfully report on the measure, are provided with each benchmark.**

- *Establish Benchmarks on Clinical Evidence:* There is a need to think about whether there are ways to determine what the optimal performance is for some of the outcome measures. The optimal performance may not be 100 percent for some measures. **If that is the case, CMS should explore setting stable benchmarks to which all physicians should be held, regardless of what the actual reporting performance rates say**. This approach would set reasonable benchmarks for all physicians to achieve and would minimize any unintended consequences of over-reporting to achieve a benchmark that is not reflective of current evidence. For example, Measure 343, Adenoma Detection Rates (ADR), we know based on evidence what the top performance rate should look like and it is roughly 25 percent for a mixed gender population. Therefore, CMS should develop benchmarks based on a maximum 25 percent given the evidence.

- *Shifting of Performance Scores from Year to Year:* Improvement scoring appears to assume that quality measure benchmarks will remain stable from year to year, when instead, the deciles will likely shift over time. Consequently, physicians may be improving their performance but this will not be captured in physician's overall points in the Quality performance category. We recognize this is the trade-off of scoring improvement on a category versus measure basis, but without more experience with the MIPS program, we are unclear how often this will happen and if it warrants a different approach. We are concerned that performance may differ across years as the number of physicians reporting on measures varies leading to potential increases or decreases in sample sizes and performance scores. In addition, differences may occur based on changes to reporting requirements across years, leading to scores that may not be reflective of true performance. For example, because of the pick-your-pace approach used in 2017, year one MIPS data may not be a representative sample of how physicians are actually performing on quality measures; yet, 2017 data will be used to create the 2020 benchmarks. While we fully support pick-your-pace as it eases the transition to MIPS, CMS needs to also take into consideration what data they will be using to form future benchmarks. **It would be useful to examine how benchmarks across years may shift to better understand whether the changes reflect more reliable data, such as a larger sample size or represent data that is insufficient for that year.**

If it is the latter, we are concerned that it could compromise CMS' ability to set a reliable benchmark for a specific year.

- *Incorporating Risk-Adjustment or Stratification*: Many of the measures were designed for a pay-for-reporting structure and not pay-for-performance. Because of the previous focus on reporting and the assumption that performance would never reach 100 percent, people have been more comfortable with outcome measures not being risk-adjusted or stratified based on patients' clinical and social risk factors. Given the shift to payment that is based on performance, there may be a greater need for developers and CMS to start risk-adjusting or stratifying at the measure level. Other challenges associated with this strategy would also need to be examined such as smaller sample sizes created as a result of either approach. For example, it is not reasonable to assume that every physician or practice could achieve an A1c less than 8 percent in every patient. The thinking when the measure was first developed was that health plans and now physicians would never achieve 100 percent. However, that approach does not work well when a program is assigning performance to deciles with an assumption that all deciles should eventually reach 100 percent. In addition, the American Diabetes Association and other medical societies are moving towards more person-centered individualized A1c goals. Engaging in shared decision making allows for patients to keep A1cs closer to eight, but not to exactly eight. Therefore, if you treat more elderly or sicker patients, CMS will unfairly penalize physicians if the A1c target is eight and the top decile set at 100 percent.

  If the goal is to have a set of measures that truly align payment with the quality of care provided to patients, CMS and measure developers should be proactive and find solutions to ensure that the benchmarks represent achievable and evidence-based performance rates. We believe that risk adjustment or stratification is one solution that should be explored further.

- *High Performance Rate on a Measure:* **The AMA believes a high performance rate on a measure does not always mean that it should be removed, and there are measures for which every physician should be aiming for top performance. In addition, we question whether measures that appear to be topped out are reflective of care across all physicians**. For example, the 2015 Reporting Experience for PQRS included measures that are identified as topped out in the QPP benchmarks; yet, less than 5 percent of eligible physicians reported on some of those measures. We believe that current performance may reflect the top performers but it may not reflect true performance across all physicians. For example, when we examine the changes in rates on these measures over time, many measures demonstrated gaps in care and sufficient variation initially, however, physicians were able to improve performance across reporting periods. We are concerned that the current approach to topped out measures may discourage physicians from reporting on important aspects of care that they may not currently provide to all of their patients. These measures were deemed important to include in PQRS and now MIPS and by setting benchmarks that do not allow physicians to achieve the highest number of points, new participants will be less likely to select and report on these important measures– measures for which there is evidence that the processes can help drive improvements in patient outcomes.

V. Quality Performance Category

Cross-Cutting Measures

The AMA strongly supports CMS' continued elimination of the requirement that physicians report on cross-cutting measures or quality measures within specific domains. The AMA also supports the removal of cross-cutting measures from specialty measure sets, which will allow physicians to report on the quality measures that are most relevant to their specialty. We believe requiring the reporting of cross-cutting measures may force physicians to select and report quality measures that are not applicable to their practice.

CMS notes in the final rule that they are seeking comment on ways to incorporate cross-cutting measures into the MIPS program in the future. The AMA believes that in future MIPS program years, a physician's decision to report on cross-cutting measures should be voluntary. **The AMA opposes prescriptive requirements that force physicians to select and report cross-cutting measures.** Instead, CMS should continue to encourage physicians to report the quality measures that are most meaningful to their practice.

Topped-Out Measures

The AMA supports CMS' phased-in approach for removing "topped-out" measures from MIPS; however, we do not support CMS' proposed timeline for classifying measures as "topped out" or its proposal to cap achievement points for such measures at seven points. CMS' current strategy bases performance scores and benchmarks on data that may or may not have sufficient sample sizes and utilizes PQRS reporting rates as a starting point. PQRS had low participation rates, and it is questionable whether the numbers represent a true indication of quality. MIPS benchmarks should be developed based on MIPS reporting, not a program that sunset in 2016.

Beginning the phased-in removal of "topped-out" measures with only one year of MIPS data is also problematic due to the 2017 transition year. Because of the pick-your-pace approach used in 2017, year one data may not be a representative sample of how physicians are actually performing on quality measures. CMS has already removed a significant number of measures under MIPS, particularly measures available under the claims and EHR reporting methods, and we remain concerned that removing and capping measures too soon may lead to a gap in the measure portfolio.

The AMA does support the removal of measures when clinical evidence has changed, but we are concerned with the potential future gap that will be created by solely relying on benchmark data, without consideration of clinical factors, scientific evidence, and the importance of a measure. More research also needs to be done to determine the appropriate sample size for each quality measure before a quality measure can be determined to be "topped-out." We are concerned with CMS' blunt approach to removing "topped-out" measures, and offer the following recommendations to improve this process:

- *Process Measures:* Process measures for which there is strong evidence that fulfillment of the measure intent, such as providing or not providing a specific treatment will improve patient outcomes, should be retained. The unintended consequences of removing key "topped-out" measures are unknown. If a "topped-out" measure directly impacts outcomes and is no longer reported, could its removal cause negative effects on patient care? CMS should exercise caution in measure removal until possible unintended consequences of removing each measures have been explored.

- *Analysis:* Physician performance can vary by practice setting, patient population, geography, years in practice, volume of cases of a particular condition, or how long the physician has been reporting. CMS must examine the breadth and depth of reporting based on the number of physicians who successfully report on a measure and the length of time a measure is reported on within a given performance year.

- *Performance Results:* Performance results of a measure that is being considered for removal should be examined for any evidence of variation among subgroups defined by the above factors and other nonclinical factors.

- *Reporting Options:* CMS should not remove or classify a measure as "topped-out" until it is "topped-out" across all reporting options.

- *Data Sources:* One potential way to see if the numbers are reflecting true performance is to compare it to other current data. For example, if a study or clinical registry shows that there is still a gap in care, then the performance scores in MIPS may not reflect performance across all physicians. The results of these subgroup analyses should also be shared with the relevant stakeholders.

- *Small Sample Size:* CMS has a history of removing measures that have low reporting rates without necessarily considering the specialties that might be reporting them. For example, some measures may only be reported by a small number of clinicians, such as pediatric specialists, and yet that small number represents a significant percentage of those caring for the patients to which the measure applies.

- *Public Health:* We recommend keeping measures that track performance on major public health issues such as tobacco use and counseling, screening for alcohol use, prediabetes, hypertension, opioid use, immunizations, and hepatitis C.

- *Measures Used in Other Programs:* There are many health plan-level measures that are part of the Medicare Advantage Star Ratings system that are reliant on clinical action. To ensure compliance, the private plans incorporate them into physician contracts. For purposes of alignment, CMS should consider alignment across other programs when deciding whether to remove or retain certain measures.

Furthermore, CMS should not penalize physicians for reporting on "topped out" measures by capping the number of achievement points at seven points. Physicians should be eligible to earn maximum achievement points for reporting such measures until a measure is removed. Capping achievement points adds to the complexity of scoring and ignores that there are multiple factors that go into the decisions physicians make for reporting on specific measures. It also ignores that CMS is making classifications on measures based on extremely faulty data with low reporting rates.

<u>QCDR measures</u>
Many medical specialty societies are developing tools such as Qualified Clinical Data Registries (QCDRs) to help physicians incorporate systems of learning into their practice to improve quality of care, provider workflow, patient safety, and efficiency. Capturing data through a registry allows for its collection and tracking across settings and disease states including, inpatient versus outpatient settings,

acute episodes versus chronic disease, surgical versus nonsurgical interventions, and resource-intensive versus relatively inexpensive therapies. However, for the improvements to be made quality measurement must move beyond snapshots of care which focus on random individual measures to a learning system with a broad focus. Utilizing specialty-led QCDRs provides an opportunity to evaluate care within an entire specialty, as well as at the individual physician level.

Moving to align standards for reviewing QCDR measures with the Call for Measures process would put the QCDR pathway in jeopardy. The current Call for Measures process does not recognize the uniqueness of QCDRs, evaluates measures on an individual basis, and is not agile. The Call for Measures process requires stewards to put their measures with testing data before CMS a year and a half prior to the start of the reporting period. For example, if a QCDR steward was interested in having their measure(s) in the 2019 QPP they would have had to submit the measure(s) and required materials to CMS in June of 2017.

We remind CMS that the statutory intent of Section 1848(q)(2)(D)(vi) of MACRA is to provide flexibility for QCDRs. Based on statue, QCDRs are not subject to certain requirements, such as inclusion of measures on the annual list of quality measures, publication in peer-reviewed journals, and endorsement by a consensus based entity. In addition, QCDR measures are exempt from consideration of whether measures address measure gaps and the priority given to outcome, patient experience, care coordination, and appropriate use measures. This reflects the intent to allow specialties to develop and select QCDR measures outside the prescriptive process used to develop and choose general quality reporting measures.

To improve the QCDR process, CMS must recognize that changes to QCDRs, registries or EHRs require significant financial resources and time to plan, incorporate, and test. This time-lag limitation becomes very challenging when CMS makes annual changes to quality requirements or technology functionality. In addition, changing the QCDRs process and expectations of QCDRs on a yearly basis creates the perception among specialty-led QCDRs that the changes are arbitrary and lack evidence or reason. The annual changes are also administratively burdensome and do not allow sufficient time for implementation. **Therefore, there must be consistency from year to year. As highlighted in our 2018 QPP comments, it is unrealistic to expect that changes can be easily adopted by the start of the performance period when sponsors of QCDRs often only learn of the changes during the annual CMS QCDR "deeming" process or proposed rule.**

We offer the following suggestions to improve the process:

- Assign a single point of contact to each QCDR to reduce communication breakdowns and conflicting messaging.
- Set up a review process where CMS and its contractor consult with appropriate physician experts and QCDR stewards to ensure sufficient clinical expert review on the importance and relevancy of a measure. One entity suited to do this is the National Quality Registry Network (NQRN) through the PCPI, of which the majority of specialty society QCDR stewards are members. Importantly, PCPI membership, and participation in NQRN, is open to a broad range of health care industry stakeholders who contribute their diverse and well-informed perspectives to the QCDR review process. The NQRN is a network of individuals affiliated with PCPI member organizations that are operating, planning, or otherwise interested in registries; using information from registries to improve patient outcomes; and providing technology and infrastructure such as

registry platforms and data standards. The PCPI QCDR committee is another forum for addressing common issues.

- Set up a system to properly record and track ownership rights, including making ownership information CMS collects available to QCDRs to better facilitate sharing of QCDR measures between QCDR stewards.
- QCDR self-nomination application and materials should be updated to outline all of the information needed to determine QCDR status to avoid delays and misunderstandings.
- Provide at least a 60 day notice of any changes to the QCDR vetting process, including review of measures and a minimum of 30 days to appeal changes.
- If a measure is provisionally approved and CMS must receive data, allow for the QCDR to collect such data for at least one full year. Therefore, measures provisionally approved for the 2017 performance period would be submitted with the 2019 self-nomination application.

We urge CMS to work with specialty-led QCDR stewards to further improve the process and ensure a viable and private sector-run innovative reporting option. If changes are not made, many specialty QCDRs have stated they may not continue to seek QCDR status because of the escalating administrative burden and arbitrary nature of the vetting process that often lacks evidence and operates on unrealistic timelines.

VI. APM CEHRT Requirement

In the 2018 Final QPP rule, CMS seeks comment on whether it should consider revising the 50 percent CEHRT requirement for Advanced APMs. The AMA urges CMS to only require physicians to use CEHRT as part of an APM if CMS has verified that CEHRT can deliver the information needed by participants in the APM efficiently and at an affordable cost.

Too many physicians have found that instead of helping them to deliver higher-quality, more coordinated care, CEHRT reduces the time they can spend with their patients and increases administrative costs. Therefore, in future years CMS should only require the use of CEHRT if it has the functionality needed for the APM. APMs should be encouraged to leverage technology to support the goals of the APM and help participants improve communication, patient engagement, collaboration, diagnosis, treatment planning, and quality. While CEHRT supports some basic, fundamental functionality to help APMs achieve their goals, the AMA has heard from physicians that CEHRT often needs to be enhanced or supplemented before it is truly useful. Any CEHRT requirement must also provide sufficient flexibility to recognize custom functionality that "builds on" CEHRT – a concept taken directly from one of CMS' priorities for new ACI measures in the MIPS program.

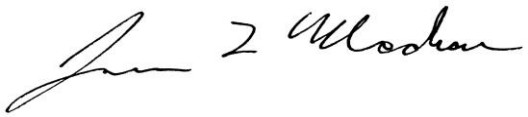
Often these additional enhancements are layered on top of CEHRT, creating a new and improved experience for patients and their care teams, and a greater return on investment than the original purchase of the EHR. The effort and value of refining the EHR experience based on patient and physician need should be rewarded by CMS. Going forward, CMS and the Office of the National Coordinator for Health Information Technology should take greater responsibility for proactively ensuring that CEHRT includes the capabilities physicians need to enter, retrieve, share, and analyze clinical data in APMs.

VII. Automatic Extreme and Uncontrollable Circumstance Policy

The AMA strongly supports the establishment of an automatic extreme and uncontrollable circumstance policy for the Quality Improvement Activities, and Advancing Care Information performance categories for the 2017 MIPS performance period. We appreciate CMS' recognition of the AMA's advocacy efforts to ensure physicians affected by natural disasters such as hurricanes and wildfires will avoid a MIPS penalty. We agree CMS should reweight the scores of physicians affected by an extreme and uncontrollable circumstance to zero percent of the final score, resulting in a final score equal to the performance threshold, unless the physician chooses to submit MIPS data. The AMA will continue to work with CMS in the future to ensure physicians are not penalized for circumstances outside of their control.

We thank you for the opportunity to provide input on this final rule and look forward to continuing to work with CMS to ensure that MIPS and APMs realize their potential to support the ongoing transformation of health care delivery. If you have any questions regarding this letter, please contact Margaret Garikes, Vice President of Federal Affairs, at margaret.garikes@ama-assn.org or 202-789-7409.

Sincerely,



James L. Madara, MD

Attachment

MIPS Benchmarking Exploration

November-December, 2017

Contents

Executive Summary

In an attempt to examine and explore the current MIPS evaluation methodology, we compared physician performance under the current MIPS methodology to performance as determined by the 5-Star Rating methodology for several quality measures. The 5-star ratings use the "ABC" methodology, whereas MIPS calculates deciles of submitted measures when they meet certain criteria (e.g., a reporting rate of at least 50%, there are at least 20 patients from at least 20 submitting entities, etc.). Therefore, the 5-star ratings have 5 levels of quality and MIPS creates 9 levels of quality (e.g., deciles 10 through 3 and "less than 3"). Using the MIPS methodology, topped-out measures may result in fewer than 9 levels, since there may not be 10 fully-defined deciles. For example, if 50% of physicians score 100%, then the 6th through 10th deciles are the same, and the MIPS deciles will appear for the top decile (the 10th, namely the 50% who scored 100%), and for deciles 5, 4, 3, and less than 3.

Our goal was to compare performance under these methodologies for a variety of measures, including those that are topped out.

Through our examination, we found that the two methodologies (MIPS and 5-star) resulted in inconsistent ratings and comparisons. In several instances, physicians deemed to be of similar quality by one methodology were classified as having different levels of quality by the other methodology. Additionally, some physicians classified in the highest (or lowest) level of quality by one methodology were not classified as such by the other methodology. This analysis does not attempt to determine which methodology more accurately reflects true quality or true differences in quality. However, the fact that the two methodologies produce different results when rating and ranking the same physicians implies that at least one of the methodologies is lacking, and suggests that further thought and testing is necessary. If MIPS proceeds with the current methodology,

the inconsistencies demonstrated here could result in physician frustration and dissatisfaction, and could ultimately lead to a lack of confidence in the MIPS program. Further, these inconsistencies would also send mixed signals to patients who might make incorrect assumptions about physician quality when deciding where to seek care.

Benchmark Analysis

Objective: to compare individual physician performance when evaluated using both the 5-star ratings and the MIPS deciles.

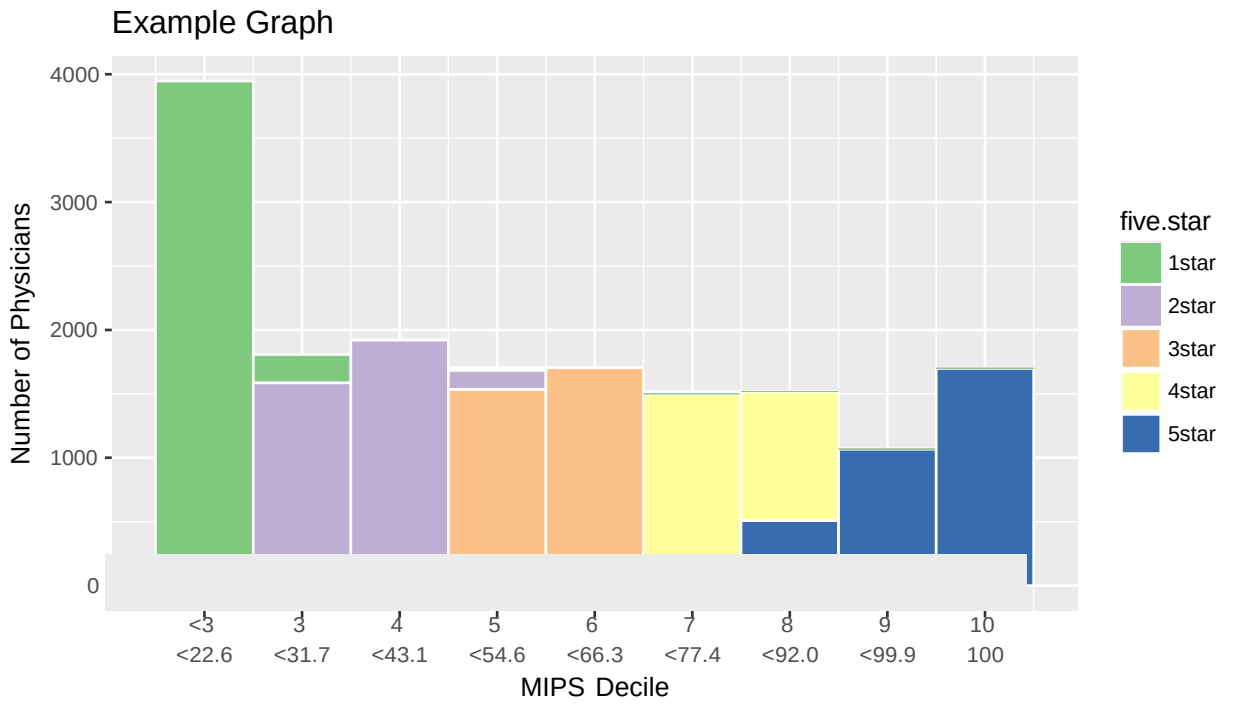
If both methods for comparing physicians are reasonable, then they should produce relatively consistent results. If, on the other hand, there are physicians who score well under one method and poorly under another, then that implies that some improvements can be made to one or both methods.

How to read the graphs:

For each physician, we classify them into a MIPS decile AND identify their 5-star rating. Across the bottom axis are the MIPS deciles, and the height of each bar reflects the number of physicians in that decile. The COLORS within each bar indicate the 5-star ratings of the physicians in that bar.

So, for example, in the example graph below, decile 7 represents a performance of < 77.4 but greater or equal to 66.3, and decile 8 represents a performance of <92.0 but greater or equal to 77.4. The number of physicians in deciles 7 and 8 looks to be about 1500 for each. In decile 7, the bar is entirely yellow, which means everyone in this decile is also a 4-star physician. In contrast, the bar in decile 8 is made up of both yellow and blue, which means there is a mix of 4-star and 5-star physicians in decile 8.

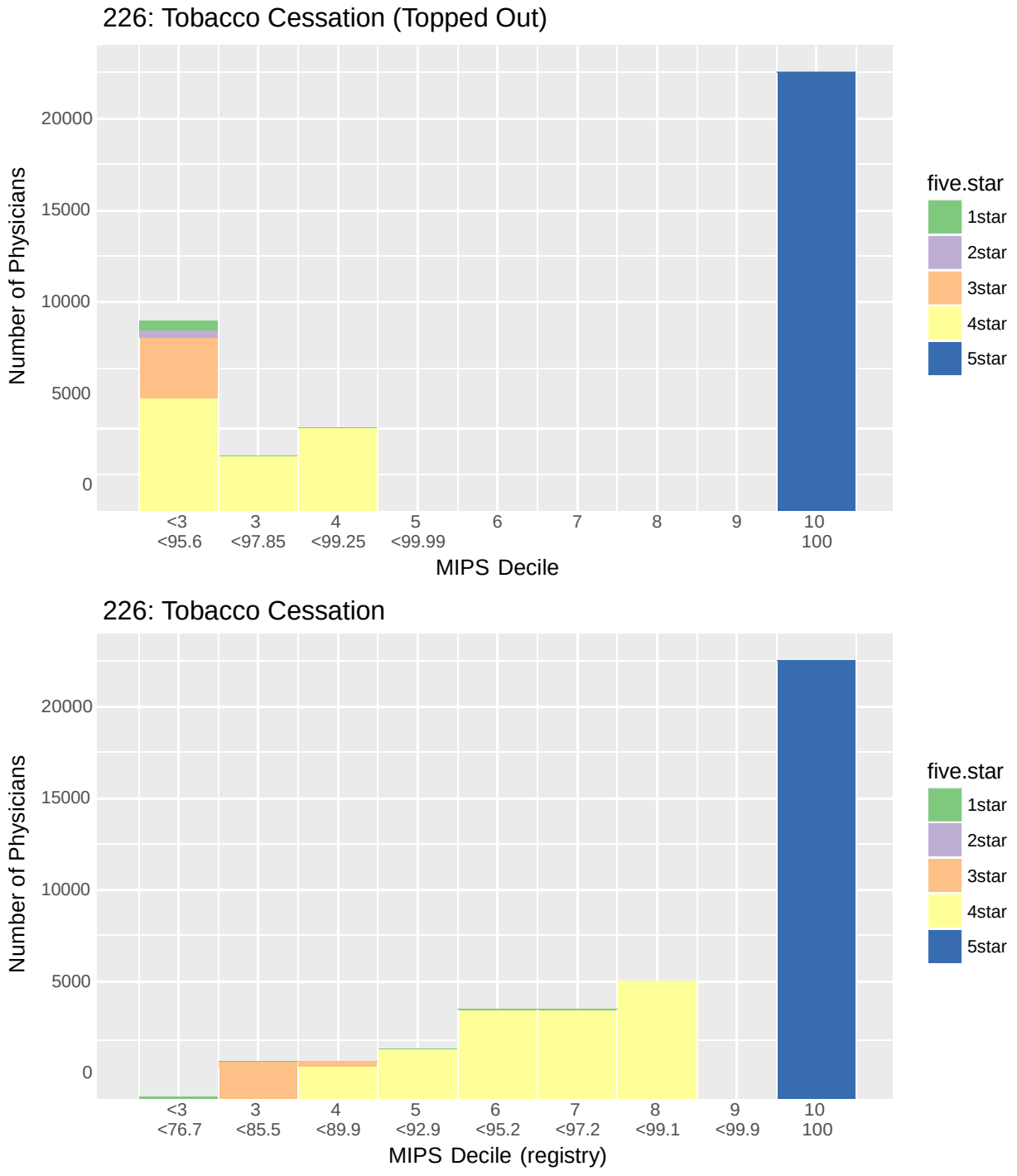
These graphs can reveal when the two methodologies (5-star vs MIPS) classify the same physician(s) differently. For example, if a single color is spread over many deciles, that suggests that among physicians who the 5-star method rates all as roughly of equivalent quality, the MIPS methods suggests that they actually differ in quality. The opposite can also be identified: if within a single decile (which contains physicians the MIPS method considers to be of similar quality) there are multiple colors, then the 5-star rating considers those physicians to be of differing quality. This inconsistency can potentially result in a lack of faith in either method.



Other notes:

1. The data represents only the "individual" physician dataset (which is the large majority of all of the data).
2. There are some MIPS measures that have multiple sets of deciles: one using claims, one using EHR data, and one using registry data. Unless otherwise noted, in these instances we used the claims-based deciles.

Measure 226

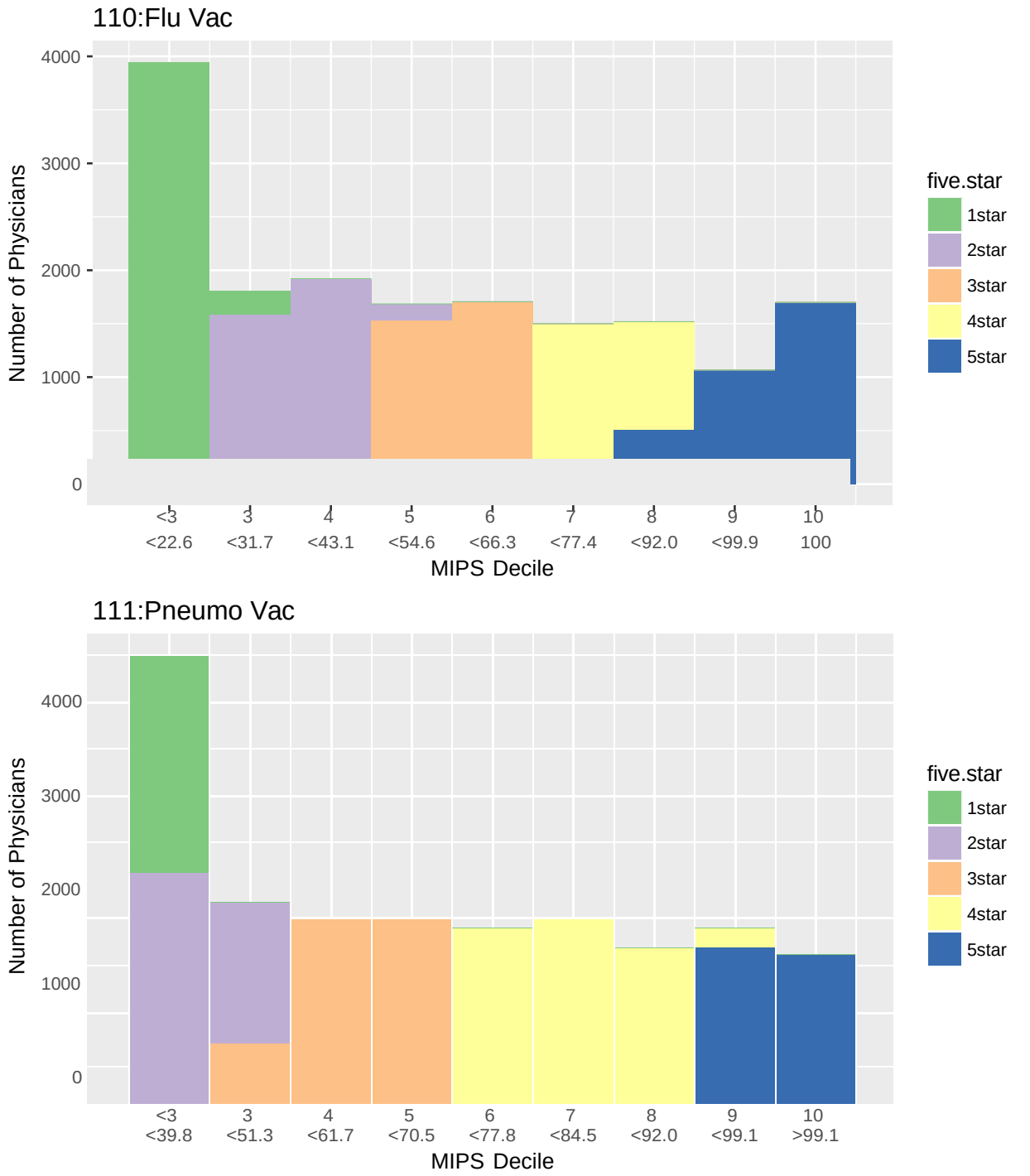


Measure 226: This is a topped out measure for MIPS when using the claims-based deciles (top graph). For this reason, performance even slightly less than 100% results in physicians landing in one of the lowest deciles. Intuitively, one might think that a performance of 95% (i.e., the lowest decile) shouldn't really be considered that much worse than a performance of 100%. And, in fact, we can see that almost EVERY physician in those lowest deciles are, in fact, of relatively high quality according to the Five Star Rating methodology (all

of the yellow represents 4-star facilities).

When using the registry-based MIPS deciles, the measure isn't topped out, but 4-star facilities (yellow), all of which the 5-star rating suggests are of similar quality, are of varying quality according to the MIPS methodology, as evidenced by their appearance in deciles 4 through 8.

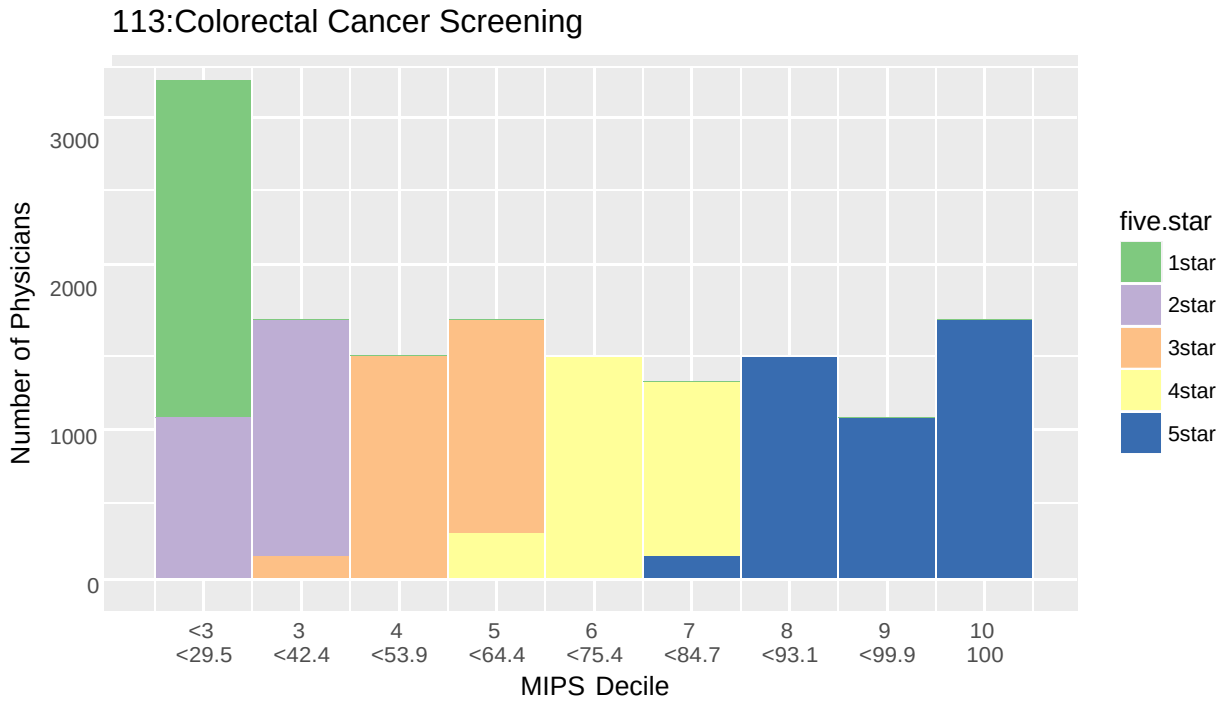
Measures 110 and 111



Measures 110 and 111: These graphs demonstrate a situation where physicians deemed to be of roughly similar quality by the Five Star method (e.g., 3-star and 4-star physicians in orange and yellow) are spread over a wide range of MIPS deciles (decile 3 to 9), suggesting that by the MIPS methodology there is quite a difference in quality in those physicians. Which are we to believe? For measure 111, when comparing a physician with a performance of 99.0 to one with a performance of 51.2, the 5-star rating suggests that

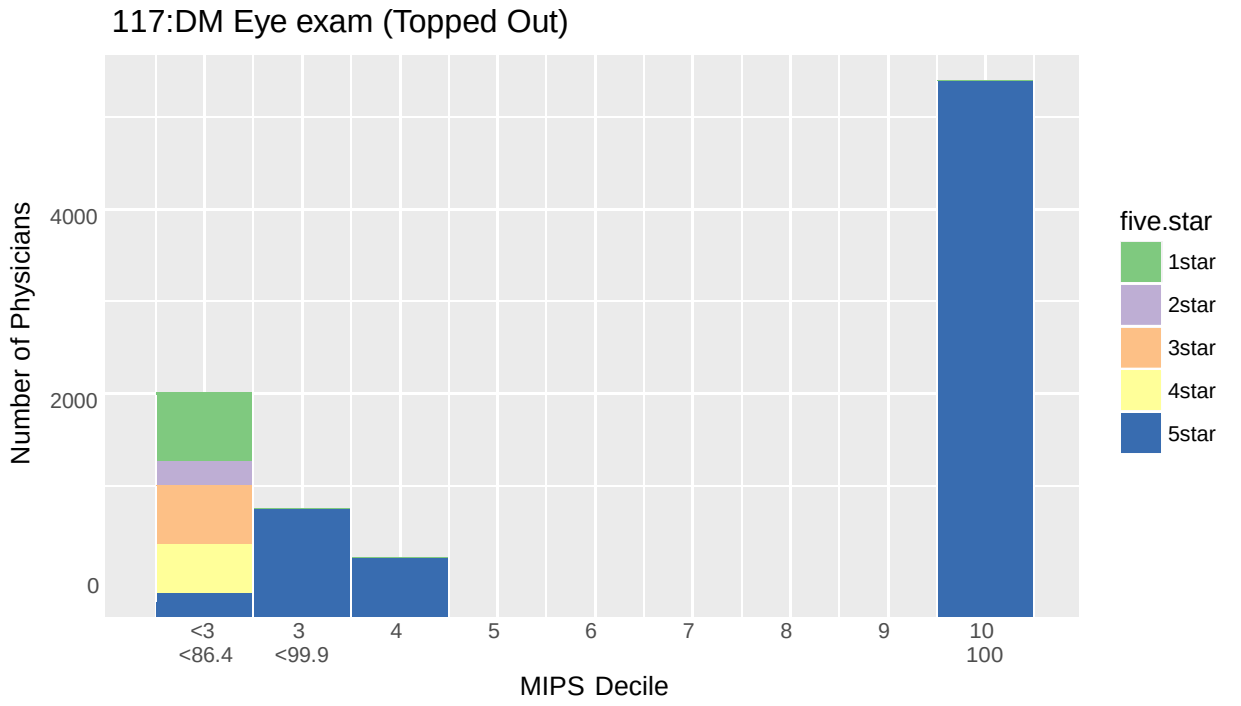
the physician with a performance of 99.0 is only 1 star better (4-star vs 3-star), while MIPS indicates that physician is 6 deciles higher.

Measure 113



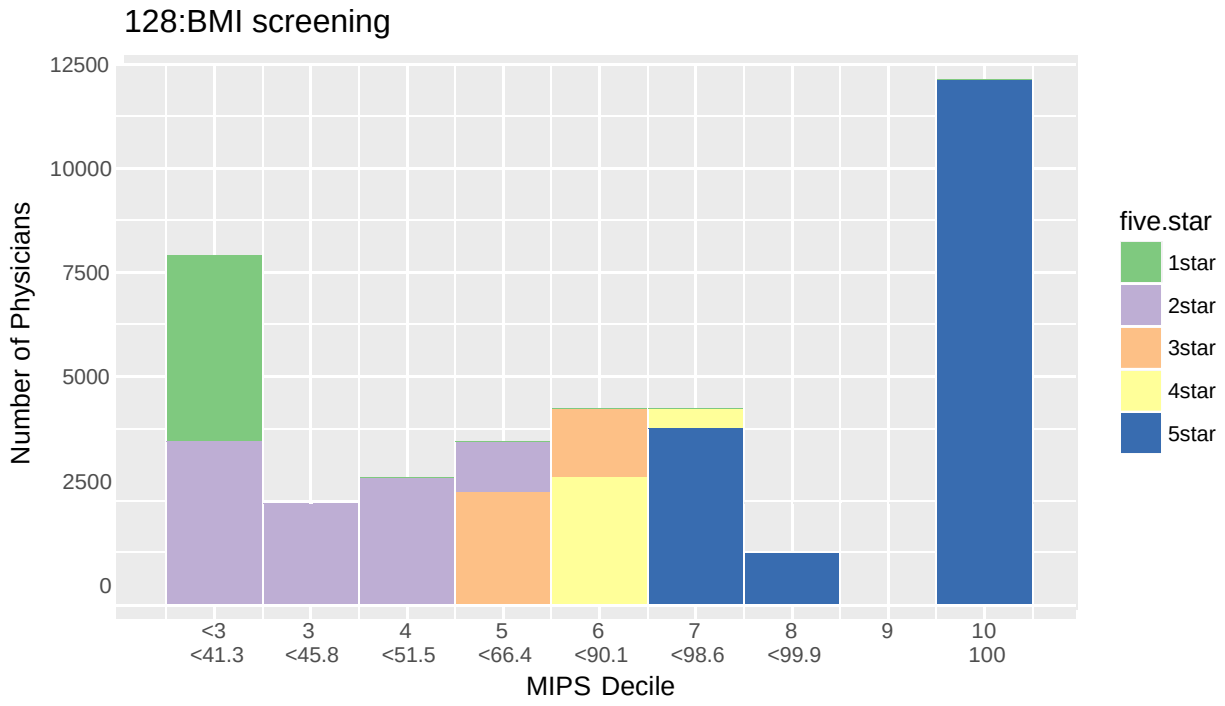
Measure 113: We have a similar inconsistency with this measure. Some 5-star physicians drop as low as decile 7 and some 4-star physicians drop to decile 5. Is everyone in the blue roughly the same quality (which is what the 5-star rating says) or are there noticeably different levels of quality (given that the MIPS method spreads them across 4 deciles)?

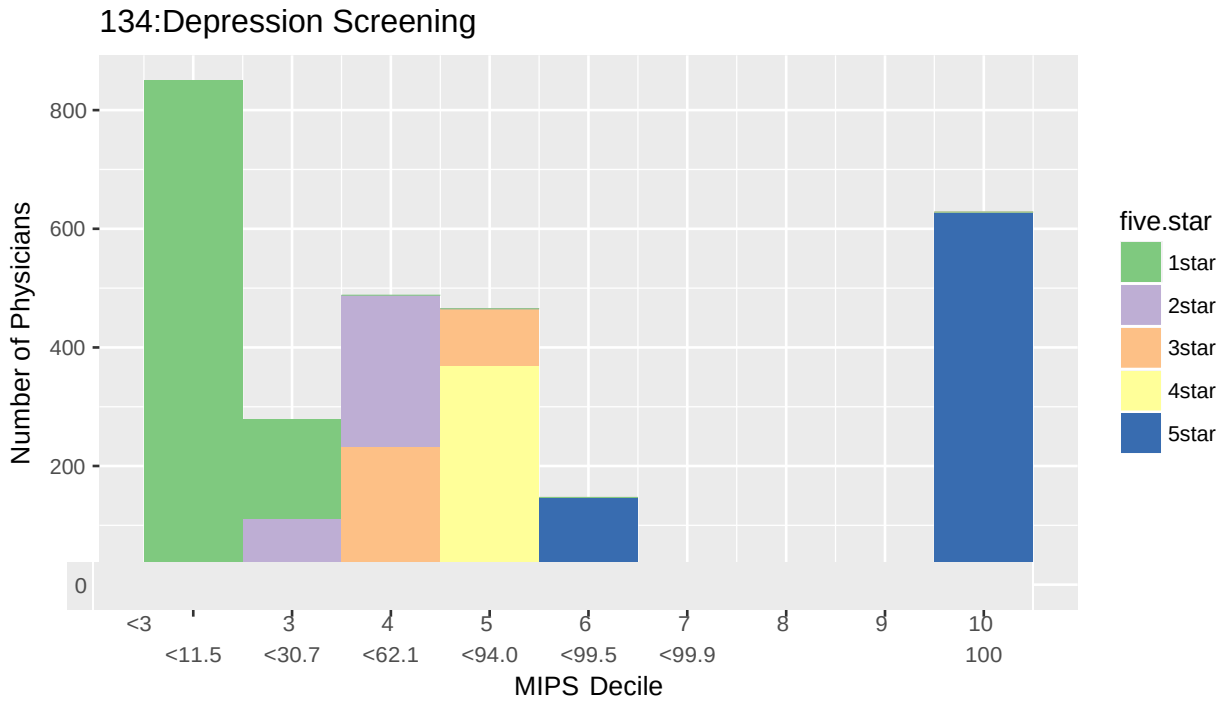
Measure 117

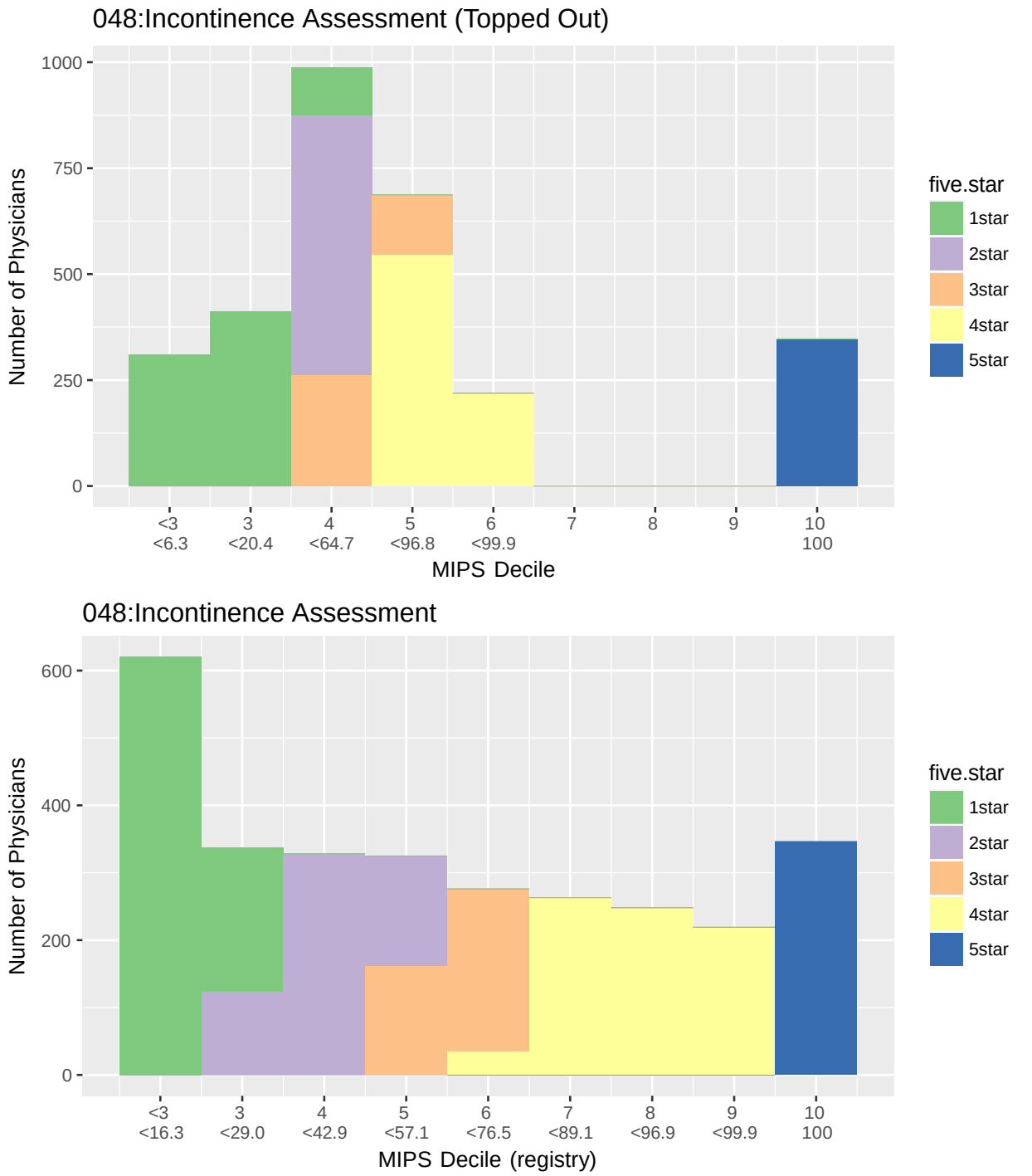


Measure 117: Here we have another topped out measure and it demonstrates both types of inconsistencies: 5-star physicians are spread across the entire spectrum of MIPS deciles (5-star says they are similar, MIPS says they couldn't be more different), and within the lowest decile we have all levels of the 5-star rating (5-star says they are different, but MIPS says they are roughly the same).

Measures 128, 134, and 48







Measures 128, 134, and 048: Again, these demonstrate a lot of inconsistencies: 5-star physicians may all be in decile 10 (Measure 48), or they may drop as low as decile 5 (Measure 134), depending on the measure. Similarly, there are 1- and 2-star physicians as high as deciles 4 and 5. And the fact that there are several colors bunched within 1 or 2 deciles of one another reflects physicians that MIPS says are roughly similar but the 5-star method says are quite different.