NATIONAL QUALITY FORUM
Driving measurable health improvements together

Memo

October 1, 2020

To: Cost & Efficiency Standing Committee

From: NQF staff

Re: Post-comment web meeting to discuss public comments received and NQF member expression of support

Introduction

NQF closed the public commenting period on the measures submitted for endorsement consideration to the Spring 2020 measure review cycle on September 14, 2020.

Purpose of the Call

The Cost & Efficiency Standing Committee will meet via conference call on October 1, 2020 from 3:00pm to 5:00pm ET. The purpose of this call is to:

- Review and discuss comments received during the post-evaluation public and member comment period;
- Provide input on proposed responses to the post-evaluation comments;
- Review and discuss NQF members' expression of support of the measures under consideration; and
- Determine whether reconsideration of any measures or other courses of action are warranted.

Standing Committee Actions

1. Review this briefing memo and draft report.
2. Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see comment table and additional documents included with the call materials).
3. Review the NQF members' expressions of support of the submitted measures.
4. Be prepared to provide feedback and input on proposed post-evaluation comment responses.

Conference Call Information

Please use the following information to access the conference call line and webinar:

Speaker dial-in #:1-800-768-2983
Access Code: 7445915
Web Link: https://core.callinfo.com/callme/?ap=8007682983&ac=7445915&role=p&mode=adt Comment Web Meeting

Background

Identifying and providing incentives for providers to deliver efficient care (i.e., high quality, lower cost) requires quality measures as well as cost and resource use measures. Such measures position the healthcare system to evaluate the efficiency of care and stimulate changes in practice to improve efficiency. The Cost and Efficiency Portfolio Standing Committee oversees NQF's portfolio of cost and efficiency measures. Measures in this portfolio address cost of care measures, which calculate total healthcare spending. The

portfolio also includes efficiency measures, which NQF defines as the resource use or cost associated with a specific level of performance with respect to the other five Institute of Medicine (IOM) aims of quality: safety, timeliness, effectiveness, equity, and patient-centeredness.

During the web meeting held on July 10, 2020, the 24 members of the Cost and Efficiency Standing Committee evaluated six newly submitted measures. The Standing Committee recommended one measure for endorsement. The measure recommended for endorsement is:

- NQF 3562 Medicare Spending Per Beneficiary – Post Acute Care Measure for Long-Term Care Hospitals (Acumen, LLC)

The Committee did not recommend two measures for endorsement:

- NQF 3561 Medicare Spending Per Beneficiary – Post Acute Care Measure for Inpatient Rehabilitation Facilities (Acumen, LLC)
- NQF 3574 Medicare Spending Per Beneficiary (MSBP) Clinician (Acumen, LLC)

The Committee did not reach consensus on three measures for endorsement:

- NQF 3563 Medicare Spending Per Beneficiary – Post Acute Care Measure for Skilled-Nursing Facilities (Acumen, LLC)
- NQF 3564 Medicare Spending Per Beneficiary – Post Acute Care Measure for Home Health Agencies (Abt Associates)
- NQF 3575 Total Per Capita Cost (TPCC) (Acumen, LLC)

Comments Received

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF solicits member and public comments during a 16-week comment period via an online tool on the project webpage.

Pre-evaluation Comments

NQF solicits comments prior to the evaluation of the measures via an online tool on the project webpage. For this evaluation cycle, the pre-evaluation comment period was open from August 14 till September 14 for six measures under review. NQF received 18 pre-evaluation comments. All of these pre-evaluation comments were provided to the Committee prior to the measure evaluation meeting.

Post-evaluation Comments

The draft report was posted on the project webpage for public and NQF member comment August 14 for 30 calendar days. During this commenting period, NQF received seventeen comments from six member organizations:

Member Council	# of Member Organizations Who Commented
Health Professional	5
Providers	1

We have included all comments that we received (both pre- and post-evaluation) in the comment table (excel spreadsheet) posted to the Committee SharePoint site. This comment table contains the commenter's name, comment, associated measure, topic (if applicable), and—for the post-evaluation comments—draft responses (including measure steward/developer responses) for the Committee's consideration. Please review this table in advance of the meeting and consider the individual comments received and the proposed responses to each.

Please note measure stewards/developers were asked to respond where appropriate. Where possible, NQF staff have proposed draft responses for the Committee to consider.

Comments and Their Disposition

Measure-Specific Comments

NQF 3561 Medicare Spending Per Beneficiary – Post Acute Care Measure for Inpatient Rehabilitation Facilities (Acumen, LLC)

A commenter had doubts about the value of the measure and agreed it should not be endorsed. They stated that inpatient rehabilitation facilities' funding and utilization are controversial, but they have a modest volume and impact in comparison to Skilled Nursing Facilities and Long-Term Acute Care Hospitals and more controlled utilization.

Measure Steward/Developer Response:

Thank you for your comment. CMS recognizes the need to use the MSPB-PAC measures in concert with other quality measures that are designed to capture clinical outcomes of care. These measures were developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act), in light of rising Medicare expenditures and the wide variation of Medicare spending across PAC services. The IMPACT Act specifically ensures that cost measures be considered alongside quality measures, including assessment-based ones. Additionally, the IRF providers involved in the delivery of high-quality care and appropriate discharge planning and post-treatment care coordination would be expected to perform well on these measures since beneficiaries would likely experience fewer costly adverse events. For example, our testing confirms that, on average, more efficient IRFs are associated with better discharge to community rates.

Developer Rationale for Reconsideration:

CMS, with Acumen and Abt Associates, request that the Cost and Efficiency Standing Committee: (i) reconsider its recommendation against endorsing two measures, and (ii) consider substantive issues in re-voting on three 'consensus not reached' (CNR) measures in the Spring 2020 evaluation cycle.

The developer argues that the evaluation criteria were not correctly applied for the measure and that inconsistent application of the evaluation criteria either led to a measure not being recommended for endorsement or to consensus not being reached. The full reconsideration request can be found in Appendix B.

Proposed Committee Response:

Thank you for your comment. The Committee will review and discuss them during the Post-Comment Meeting.

Action Item:

Based on comments received and the information provided by the developer, the Committee will vote to reconsider the measure.

NQF 3562 Medicare Spending Per Beneficiary – Post Acute Care Measure for Long-Term Care Hospitals (Acumen, LLC)

A commenter stated that Long-Term Acute Care Hospitals' (LTACH) funding and utilization are controversial. Though they supported the Committee's endorsement of the measure, they believed that high LTACH

utilization does not necessarily correlate with higher quality or better outcomes and suspected that there was substantial regional variation.

Measure Steward/Developer Response:
Thank you for your comment. CMS recognizes the need to use the MSPB-PAC measures in concert with other quality measures that are designed to capture clinical outcomes of care. These measures were developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act), in light of rising Medicare expenditures and the wide variation of Medicare spending across PAC services. The IMPACT Act specifically ensures that cost measures be considered alongside quality measures, including assessment-based ones. Additionally, the LTCH providers involved in the delivery of high-quality care and appropriate discharge planning and post-treatment care coordination would be expected to perform well on these measures since beneficiaries would likely experience fewer costly adverse events. For example, our testing confirms that, on average, more efficient LTCHs are associated with better discharge to community rates.

Proposed Committee Response:
Thank you for your comment. The Committee will review and discuss them during the Post-Comment Meeting.

Action Item:
The Committee should review the comment received and provide a proposed response.

NQF 3563 Medicare Spending Per Beneficiary – Post Acute Care Measure for Skilled-Nursing Facilities (Acumen, LLC)

A commenter expressed nonsupport for the measure, as they stated post-acute SNF utilization is not necessarily meaningful in and of itself.

Measure Steward/Developer Response:
Thank you for your comment. CMS recognizes the need to use the MSPB-PAC measures in concert with other quality measures that are designed to capture clinical outcomes of care. These measures were developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act), in light of rising Medicare expenditures and the wide variation of Medicare spending across PAC services. The IMPACT Act specifically ensures that cost measures be considered alongside quality measures, including assessment-based ones. Additionally, the SNF providers involved in the delivery of high-quality care and appropriate discharge planning and post-treatment care coordination would be expected to perform well on these measures since beneficiaries would likely experience fewer costly adverse events.

Proposed Committee Response:
Thank you for your comment. The Committee will review and discuss them during the Post-Comment Meeting.

Action Item:
The Committee will revote on the validity criterion.

NQF 3564 Medicare Spending Per Beneficiary – Post Acute Care Measure for Home Health Agencies (Abt Associates)

A commenter expressed nonsupport for the measure, as they stated post-acute care HHA utilization is not necessarily meaningful in and of itself.

Measure Steward/Developer Response:
Thank you for your comment. CMS recognizes the need to use the MSPB-PAC measures in concert with other quality measures that are designed to capture clinical outcomes of care. These measures were developed to address the resource use domain of the Improving Medicare Post-Acute Care

Transformation Act of 2014 (IMPACT Act), in light of rising Medicare expenditures and the wide variation of Medicare spending across PAC services. The IMPACT Act specifically ensures that cost measures be considered alongside quality measures, including assessment-based ones. Additionally, the home health providers involved in the delivery of high-quality care and appropriate discharge planning and post-treatment care coordination would be expected to perform well on these measures since beneficiaries would likely experience fewer costly adverse events.

Proposed Committee Response:
Thank you for your comment. The Committee will review and discuss them during the Post-Comment Meeting.

Action Item:
The Committee will revote on reliability and validity criterion.

NQF 3574 Medicare Spending Per Beneficiary (MSBP) Clinician (Acumen, LLC)

A commenter expressed concerns with the measure specifications and reliability and attribution at the individual clinician level. They disagreed with the measure's attribution of costs to providers like primary care physicians for care they did not provide and who have limited control over many of those costs. They noted that primary care services represent a very small portion of overall costs. The commenter also had concerns about the impact of excluding patients who died on the overall model, and the lack of correlation between cost and quality measures, particularly patient outcomes. Another commenter agreed with the Committee's concerns on the scientific acceptability of the measure, expressing the need for the developer to demonstrate reliable and valid results to allow users to make meaningful distinctions in care costs. Commenters were also concerned with the lack of information on reliability results below the 25th percentile, particularly in light of the reference within the response of 2a2.3 that CMS generally considers 0.4 to be the threshold for moderate reliability and 100% of practices and clinicians with at least 20 episodes meet it.

It was stated that the higher Medicare Spending Per Beneficiary rarely correlates with better outcomes, but this is very difficult to sort out at the clinician level. A member voiced concerns about necessity of the TPCC and MSPB measures, as many of the beneficiaries captured in the episode-based measures will also be included in either or both the MSPB and TPCC measures. This would result in a beneficiary potentially being attributed to multiple providers within and across multiple measures which could magnify the impact on cost measures of any individual beneficiary and complicate differences in cost and value.

Commenters requested information and testing to demonstrate that measure's use in Merit Incentive Payment System would yield reliable and valid results and enable end users to make meaningful distinctions on the costs associated with the care provided to patients. Commenters supported the Committee's decision not to endorse this measure. They stated that outside of an ACO setting or other risk-sharing arrangement that covers all care provided to a population, the measure attributes costs to providers for care they did not provide and who have limited control over many of those costs. Concerns were shared that the measure did not provide insight into which treatments were most effective in providing high quality, low cost care. Episode-based cost measures were brought up as a better approach to evaluating value. It was also recommended radiation therapy be excluded from post-trigger inpatient and outpatient components.

Measure Steward/Developer Response:
The developer thanks the commenters for their comments. Their responses are structured to combine comments on similar topics. Many of these comments were raised by other commenters and addressed during the Standing Committee evaluation meeting. The developer has focused on new points and briefly recapped where issues have already been discussed. They also refer to their Request for Reconsideration included in Appendix B for their overall responses to the concerns related to the reliability, face validity, and empirical validity of the measure. Please see their full list of responses in Appendix C, as well as the comment table.

Developer Rationale for Reconsideration:

CMS, with Acumen and Abt Associates, request that the Cost and Efficiency Standing Committee: (i) reconsider its recommendation against endorsing two measures, and (ii) consider substantive issues in re-voting on three 'consensus not reached' (CNR) measures in the Spring 2020 evaluation cycle.

The developer argues that the evaluation criteria were not correctly applied for the measure and that inconsistent application of the evaluation criteria either led to a measure not being recommended for endorsement or to consensus not being reached. The full reconsideration request can be found in Appendix B.

Proposed Committee Response:

Thank you for your comment. The Committee will review and discuss them during the Post-Comment Meeting.

Action Item:

Based on comments received and the information provided by the developer, the Committee will vote to reconsider the measure.

NQF 3575 Total Per Capita Cost (TPCC) (Acumen, LLC)

Similar concerns to NQF #3574 were raised by commenters for this measure regarding measure specification, attribution at the individual clinician level, rare correlation with better outcomes, exclusion of patients who died in the overall model, the lack of correlation between cost and quality measures, and scientific acceptability. The commenters also mentioned that they were unsure the developer showed that the measure correlates to any one quality measure within the MIPS program and requested the Committee discuss whether the results of the attribution and validity in the measure could lead to negative unintended consequences. They were also concerned with the lack of information on reliability results below the 25th percentile, particularly in light of the reference within the response of 2a2.3 that CMS generally considers 0.4 to be the threshold for moderate reliability and 100% of practices and clinicians with at least 20 episodes meet it.

A commenter stated the attribution methodology is a significant threat to the validity of the measure. It was acknowledged that the TPCC eliminates the problem of attributing costs that occurred before the clinician ever saw the patient. However, the current approach could attribute the measure to practices and clinicians that billed E&M claims lower than desirable percentages. There were concerns that the attribution methodology assumes that a primary care relationship exists if two things happen within three days or three months, and not otherwise. This would lead to significant problems when considering best practices in care. In addition, an oncologist will not know if they qualify for the TPCC measure, as the exemption is applied retrospectively based on a measurement of candidate events for which the oncologist bills for chemotherapy or radiation therapy services.

A commenter also stated that within the attribution methodology, there is not an end to the clinician's primary care responsibility for the patient when a Medicare beneficiary switches to a new clinician. TPCC assigns responsibility for all Medicare Part A and B costs for 12 months after attribution. This would result in attribution to multiple clinicians, as patients switch providers. This would be inappropriate as only one clinician would be coordinating the patient's care and the other will not be aware of any services provided.

There was a request that all medical and radiation oncologists be excluded from the TPCC measure. It was recommended radiation therapy be excluded from post-trigger inpatient and outpatient components.

Commenters believed that the concerns outlined by the Committee during the initial review along with deficiencies in the attribution methodology should result in the measure not achieving a recommendation for endorsement. It was, overall, urged that the Committee should not endorse this measure.

Measure Steward/Developer Response:

The developer thanks the commenters for their comments. Their responses are structured to combine comments on similar topics. Many of these comments were raised by other commenters and addressed during the Standing Committee evaluation meeting. The developer has focused on new points and briefly recapped where issues have already been discussed. They also refer to their Request for Reconsideration included in Appendix B for their overall responses to the concerns related to the reliability, face validity, and empirical validity of the measure. Please see their full list of responses in Appendix C, as well as the comment table.

Proposed Committee Response:

Thank you for your comments. The Committee will review and discuss them during the Post-Comment Meeting.

Action Item:

The Committee will revote on the validity criteria.

NQF Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ("support" or "do not support") for each measure submitted for endorsement consideration to inform the Committee's recommendations. One NQF member provided their expressions of support: See Appendix A.

Appendix A: NQF Member Expression of Support Results

One NQF members provided their expressions of nonsupport. Two of six measures under consideration received nonsupport from NQF members. Results for each measure are provided below.

3574: Medicare Spending Per Beneficiary (MSPB) Clinician (Acumen LLC/CMS)

Member Council	Support	Do Not Support	Total
Health Professional		4	4

3575: Total Per Capita Cost (TPCC) (Acumen LLC/CMS)

Member Council	Support	Do Not Support	Total
Health Professional		3	3

Appendix B: Reconsideration Request & Summary

NQF Request for Reconsideration Summary

CMS (the steward), with Acumen and Abt Associates (the developers), have requested that the Committee (i) reconsider their recommendation against endorsing #3561 and #3574, and (ii) consider key points for the correct application of evaluation criteria in re-voting on #3563, #3563, and #3575. In the request, the steward and developers responded to the Committee's concerns and provided further clarification and referenced NQF guidance on the evaluation of the reliability and validity criteria. The steward and developers have addressed overarching issues for the first measure when they were discussed, in line with the Committee's approach during the evaluation meeting. These points should, however, be considered across all relevant measures.

The steward and developers clarified the applicable standards for face validity based on materials documenting NQF evaluation criteria. NQF has stated that face validity – while the lowest form of validity – is sufficient for new measure submissions. As the Committee did not express concerns about the systematic assessment of face validity from recognized experts, the correct application of this criterion should result in measures #3574 and #3575 passing validity. In addition, the steward and developers provided responses to the Committee's questions around attribution for #3574 to clarify that revised methodology captures team-based care by allowing more than one clinician to be attributed an episode.

The steward and developers clarified key issues related to the risk adjustment models for the evaluation of empirical validity. They noted that the Committee's discussion of payment variables and social risk factors did not take into consideration all the factors that NQF has outlined in their evaluation guidance on identifying variables for use in a risk adjustment. Specifically, the use of some prospective payment system variables, such as functional/cognitive status, for post-acute care settings would not meet NQF guidance for considering risk-adjustment variables because the link between excessive spending in the IRF, HH, and SNF settings and excessive use of therapy and inappropriate coding of patient status on assessment instruments has been well documented, including by MedPAC (2016, 2019). Therefore, inclusion of these variables would violate NQF guidance that variables be resistant to gaming, are not indicators of the care provided, and be present at the start of care.

Further, the steward and developers noted that the Committee's concerns about not including social risk factors in risk adjustment appeared to focus only on reducing bias in measurement, without considering the testing results and the NQF's guidance about being careful not to mask disparities in care. The steward and developers pointed out that adjusting for SRFs may, in fact, mask disparities in care, creating a lower standard of care for beneficiaries with higher social risk; and that empirical testing recommended by NQF (2014) reveals that poorer performance for high social risk individuals is closely tied to providers themselves, rather than individual beneficiaries. Including SRFs would also penalize providers in some settings/models for taking on beneficiaries with high social risk (i.e., it would exacerbate bias in measurement, not reduce it), due to a negative relationship between social risk and spending in those settings. Lastly, the steward and developers pointed out that the impact on provider scores of not including SRFs in risk adjustment was carefully considered and was found to be minimal.

The steward and developers also responded to the Committee's concerns about low reliability for a subset of small or lower-volume providers at the 25th percentile, pointing out that all measures passed the SMP. NQF has endorsed measures, as recently as 2020, with overall mean reliability substantially lower than these values.

Lastly, the steward and developers addressed the Committee's concerns about low R-squared values by clarifying the interpretation and summarizing testing results that show that the risk adjustment model is functioning well overall. The steward and developers noted that the values should be evaluated in a broader

context, as NQF has done in the past when it endorsed measures with similar R-squared values. In this case, the broader context includes consideration of the measures' goals and the role of clinically unrelated services in the measures' ability to differentiate provider performance.

TO: Cost and Efficiency Standing Committee, National Quality Forum

FROM: The Centers for Medicare & Medicaid Services, Acumen, LLC, Abt Associates

DATE: September 14, 2020

SUBJECT: Spring 2020 Cost Measures: Request for Reconsideration

CMS, with Acumen and Abt Associates, request that the Cost and Efficiency Standing Committee: (i) reconsider its recommendation against endorsing two measures, and (ii) consider substantive issues in re-voting on three 'consensus not reached' (CNR) measures in the Spring 2020 evaluation cycle. The Standing Committee's current recommendations are summarized in Table 1, below:

Table 1. Cost and Efficiency Standing Committee Recommendations

NQF #	Measure Name	Results for Must-Pass Criteria	Standing Committee Recommendation
NQF 3561	Medicare Spending Per Beneficiary – Post-Acute Care (MSPB-PAC) Inpatient Rehabilitation Facility (IRF)	Reliability: Pass Validity CNR, Low[1]	Do not recommend
NQF 3562	MSPB-PAC Long-Term Care Hospital (LTCH)	Reliability: Pass Validity: Moderate	Recommend for endorsement
NQF 3563	MSPB-PAC Skilled Nursing Facility (SNF)	Reliability: Pass Validity: CNR	Consensus not reached
NQF 3564	MSPB-PAC Home Health (HH)	Reliability: CNR Validity: CNR	Consensus not reached
NQF 3574	MSPB Clinician	Reliability: CNR Validity: Low	Do not recommend
NQF 3575	Total Per Capita Cost (TPCC)	Reliability: Pass Validity: CNR	Consensus not reached

We discuss each measure in turn. When we believe that the evaluation criteria were not correctly applied for a particular measure, we present our reasoning for believing so and reference the applicable NQF documentation on evaluation. Inconsistent application of the evaluation criteria either led to a measure not being recommended for endorsement or to consensus not being reached.

---

[1] P's rating of 'Moderate' (Yes: 8, No: 7), then voted on the evaluation criteria and reached consensus on 'Low' (H:0, M: 5, L: 10, I:0).

#3561 MSPB-PAC IRF

We request that the Committee reconsider its overall 'do not recommend' evaluation of the MSPB-PAC IRF measure, as the must-pass validity criterion was not assessed in accordance with the correct NQF standards.

Empirical Validity

We request that the Committee reconsider its CNR/low rating, as sub-criterion 2b was not properly applied. The two main issues highlighted in the Draft Report are discussed in turn below. We believe that neither were evaluated with consideration of the correct standards.

'Alignment' of Payment Systems and Risk Adjustment

The Committee's focus on the use of prospective payment system (PPS) case-mix factors in the risk adjustment model, that factored strongly in the measure being determined as unable to capture what it was intended to capture, was inconsistent with sub-criterion 2b3. To summarize, #3562 MSPB-PAC LTCH and #3574 MSPB Clinician have risk adjustment models that include payment variables in the LTCH and Acute Inpatient Prospective Payment Systems – MS-LTC- DRGs and MS-DRGs, respectively. In contrast, #3561 MSPB-PAC IRF, #3563 MSPB-PAC SNF, and #3564 MSPB-PAC HH purposefully exclude some payment variables. As we noted during the Committee meeting, this was an explicit policy decision by the Center for Medicare (CM) and Center for Clinical Standards and Quality (CCSQ) as excessive spending in the IRF, HH, and SNF settings has historically been driven by excessive use of therapy and inappropriate coding of patient status on assessment instruments. This has been discussed extensively in the literature, including in the Medicare Payment Advisory Commission (MedPAC)'s Reports to the Congress in 2016[2] and 2019[3].

---

[2] MedPAC, Report to the Congress: Medicare Payment Policy (March 2016), http://www.medpac.gov/docs/default-source/reports/march-2016-report-to-the-congress-medicare-payment-policy.pdf. E.g.: "[T]he consistent finding that high-margin IRFs have patients who are, on average, less severely ill in the acute care hospital but more functionally disabled upon admission to the IRF suggests the possibility that coding practices contribute to greater profitability in some IRFs, especially given the comparatively low level of costs and cost growth in high-margin facilities.

Providers may differ in their assessment of patients' motor and cognitive function, resulting in payments for some IRFs that are too high relative to the costs incurred in treating their patients…. This phenomenon also would make some providers appear to be more cost-efficient than they actually are (since their costs would be lower than expected given their reported case mix)." pages 264-5

[3] MedPAC, Report to the Congress: Medicare and the Health Care Delivery System (June 2019), http://medpac.gov/docs/default-source/reports/jun19_medpac_reporttocongress_sec.pdf. E.g.: "We found large differences in the broad levels of function assigned to patients at their discharge from one setting and at their admission to the next PAC setting, and between assessment items collected for payment purposes and the uniform items used in quality reporting. Further, the differences in the functional categories favored recording function that would raise payments in [the IRF, SNF, and HH] settings and that would show larger improvement in quality performance..." page xxi

The following evaluation guidance from NQF was not followed in the Committee's review of the risk adjustment model with regards to the use of payment variables:

- NQF guidance states that developers should consider clinical, health status, and social risk factors variables for potential inclusion in a risk adjustment strategy.[4] As noted in our submitted materials and during the meeting, we did consider such variables and included a wide range of clinical factors, including IRF Rehabilitation Impairment Categories (RICs).
- NQF lists 10 factors that can indicate if variables should be considered for use in a risk adjustment strategy.[5] Of these considerations, the use of IRF, SNF, and HH functional/cognitive status payment variables would violate the following:
    - *Resistant to manipulation or gaming*. The IRF, SNF, and HH settings are highly susceptible to gaming, as providers can move beneficiaries into higher paying case-mix groups by providing excessive therapy or coding patient status.
    - *Is not an indicator or characteristic of the care provided (e.g., treatments, expertise of staff)*. The excluded IRF, SNF, and HH payment variables may be correlated with the quality of care provided, as gaming functional/cognitive status variables is correlated with providers' profitability (see MedPAC reports).
    - *Present at the start of care.* The IRF, SNF, and HH functional/cognitive status payment variables may be determined by care providers after the start of care.
    - *Clinical/conceptual relationship with the outcome of interest.* Due to the evidence of gaming, the conceptual relationship with spending is broken, as the purpose of the measures is precisely to prevent the type of excessive cost that is related to providers' own decisions and coding behavior.

Social Risk Factors in Risk Adjustment

We believe that the Committee did not appropriately apply the evaluation standards of sub-criterion 2b3 to the question of including social risk factors in the risk adjustment model. The Committee appeared to focus on reducing bias in measurement without considering other factors that NQF guidelines state should be accounted for in a risk adjustment strategy:

- *Consider patient factors (including clinical and social risk factors) that influence the measured outcome (but not factors related to disparities in care or the quality of care).* Adjusting for social risk factors may mask disparities in care, creating a lower standard of care for beneficiaries with higher social risk. This could allow for a higher rate of readmissions, complications, etc., among those with high social risk. This may be appropriate if such outcomes are outside of a provider's control, but empirical testing reveals that poorer performance for high social risk individuals is closely tied to providers themselves, rather than individual beneficiaries, with especially strong effects in particular settings. NQF recognized the importance of such testing in its 2014 technical report: "If a unit-level factor has an effect that is substantial relative to the patient-level effect, including only a patient-level covariate may result in adjustment for differences in quality of treatment."[6] Our analysis indicates that between-provider effects are larger than within-provider effects in most settings and models, as noted during the meeting. The Committee did not appropriately consider the implications of these results, and focused on the omitted variable bias.

---

[4] NQF, Committee Guidebook for the NQF Measure Endorsement Process v6.0 (Sept 2019), page 50

[5] NQF, Committee Guidebook for the NQF Measure Endorsement Process v6.0 (Sept 2019), page 50

[6] NQF, Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors (Aug 2014), page 25

- *Empirical association with the outcome of interest.* The sign of the relationship between social risk factors and MSPB-PAC scores varies by setting and model. In some cases, the sign on a social risk factor (e.g., dual status) is negative, meaning that expected costs are lower for a beneficiary with high social risk. Incorporating social risk factors into risk adjustment in these cases would then penalize providers for taking on beneficiaries with high social risk. This would have the opposite intended effect that the Committee considered which was to reduce bias in measurement.
- *Contribution of unique variation in the outcome (i.e., not redundant).* Our testing of the potential magnitude of omitting social risk factors from the MSPB-PAC IRF risk adjustment model showed a minimal impact on provider scores – this was our conclusion for each of the cost measures in this cycle. We believe that adding social risk factors to the robust risk-adjustment models is not warranted by the empirical analysis due to the minimal impact of MSPB-PAC scores, as well as the conceptual issues discussed above. For example, only 9 out of >1,000 IRFs would see a substantial reduction in their MSPB- PAC scores (>0.1SD) and only 3 would see a substantial increase.

#3562 MSPB-PAC LTCH

We do not have comments on the evaluation of this measure.

#3563 MSPB-PAC SNF

We request that the Committee consider the information summarized in the #3561 MSPB-PAC IRF section so as to apply the correct NQF standards to a re-vote on the CNR result.

#3564 MSPB-PAC HH

We provide the following comments regarding the MSPB-PAC HH measure for the Committee's consideration in re-voting on the reliability and validity criteria where consensus was not reached. We also ask the Committee to apply the information summarized in the #3561 MSPB-PAC IRF section to the re-vote.

Reliability

We believe that the application of sub-criterion 2a2 should lead to a rating of moderate for reliability. The NQF's guidance on how to evaluate reliability explicitly declines to set a minimum standard or threshold.[7] Instead, the NQF refers to 'acceptable norms'[8] and presents the question in the evaluation algorithm as whether there was 'high/moderate/low certainty or confidence that the

[7] NQF, Committee Guidebook for the NQF Measure Endorsement Process v6.0 (Sept 2019), page 45

[8] NQF, Committee Guidebook for the NQF Measure Endorsement Process v6.0 (Sept 2019), page 48

performance measure scores are reliable.'[9] Since NQF does not set thresholds, the Committee must consider all factors that help establish confidence of reliability.

We believe that the Committee should consider the following when revisiting the CNR vote on reliability:

- The literature as presented in our submission materials shows that there are many interpretations of reliability scores and that there is not a conclusive definition of what constitutes high, moderate, or low reliability.
- This measure – as with each of the cost measures in this cycle – passed the Scientific Methods Panel (SMP)'s review (H:3, M:3, L:1, I:1). As the SMP's role is to provide consistency and expertise in the scientific acceptability of measures across NQF topic areas, a departure from their recommendation should require an evidentiary burden that the SMP was incorrect or incomplete in their initial evaluation. This process was followed only for #3561 MSPB-PAC IRF and #3562 MSPB-PAC LTCH, where the Committee first voted on whether or not to uphold the SMP's rating; this process was not followed for #3564 MSPB-PAC HH or any of the other measures.
- A proxy for 'acceptable norms' in the absence of NQF standards is to consider the reliability of endorsed measures. As noted during the evaluation meeting on July 10, 2020, there are many endorsed measures, including measures that passed the Consensus Standards Approval Committee (CSAC) and the SMP in 2020, with overall mean reliability substantially below the reliability at the 25th percentile for the HH providers with the smallest case volume.[10]

Empirical Validity

We provide additional information regarding the interpretation of risk model metrics of discrimination for the Committee's consideration in re-voting on the validity criterion. The purpose of assessing validity is to determine whether the measure is capturing what it is intended to capture; however, the Committee placed undue importance on R-squared values, which – in isolation – do not answer this question.

R-squared

The Committee's interpretation of the R-squared value was that a low value indicates patient risk is not being accurately captured; however, this interpretation omits other important factors to consider in evaluating a risk adjustment model under sub-criterion 2b3.

Interpretation of R-squared Value in Cost Measure Context

[9] NQF, Committee Guidebook for the NQF Measure Endorsement Process v6.0 (Sept 2019), page 47. NQF, Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement (Sept 2019), page 24.

[10] For example, CSAC passed #0684 and #0686 in July 2020 with split-half correlations of 0.42 and 0.26, respectively. See also #0073, #0689, #3032, #2561, #0694, #0669, #0425.

The R-squared statistic describes the proportion of observed variation that is explained by the model. However, it does not determine whether the model is well-specified, unbiased, and/or appropriate for its policy goals.

To illustrate, we present an example of why some valid measures will have low R- squared values, while others will have high R-squared values. Suppose that beneficiaries were distributed uniformly over providers (i.e., every provider treated a similar sample of beneficiaries) and one-half of providers placed every beneficiary into an excessive physical therapy regime while the other half used a moderate physical therapy regime. In this case, differences in observed cost are not related to beneficiary characteristics and are instead intimately tied to provider choices. As such, the R-squared of a risk adjustment model will be low because beneficiary characteristics do little, if anything, to explain observed cost.

The measure in this example would, however, still be valid; in fact, the low R-squared illustrates the need for a measure as there is substantial cost variation that is not reasonably related to beneficiary characteristics, and the measure can distinguish providers who are undertaking excessively high cost. This example captures a core reason for why #3564 MSPB- PAC HH and #3563 MSPB-PAC SNF measures capture the concept of cost that is intended to be captured, and happen to have low R-squared values.

Role of Clinically Unrelated Services in Interpreting R-squared Values

Unrelated services – such as planned hospital admissions and routine management of certain preexisting chronic conditions – were purposefully and carefully excluded to improve the ability to interpret and compare MSPB-PAC scores across providers. Since unrelated services may be well predicted by patient risk factors, excluding them can reduce the explained portion of the cost variance and the model's adjusted R-squared.

Consider the example of beneficiaries with end-stage renal disease (ESRD). If a cost measure were to include the costs of routine dialysis, the R-squared would be higher than if that measure excluded the costs of dialysis. This is because the ESRD indicator variable in the risk adjustment model would explain much of the cost variation due to dialysis. However, if routine dialysis is not prescribed by or within the scope of home health providers, the measure that includes dialysis costs may have a higher R-squared but would be less valid, as it is capturing cost that a provider cannot influence. On the other hand, the cost measure that excludes the cost of routine dialysis may have a lower R-squared but the variation that is being captured is used to distinguish between good and poor performance.

R-squared Must be Evaluated Alongside Other Risk Adjustment Model Testing

We ask the Committee to review the extensive testing materials on the validity of the risk adjustment models contained in our measure submission. For convenience, we summarize key points here:

- The models control for over 100 comorbidities (including comorbid interactions), case mix categories, and patient risk factors.
- Extensive clinical review was performed by clinicians with experience providing care in PAC settings, in collaboration with Medical Officers at CMS, to identify and review relevant risk factors.

- The model includes consideration of the policy and practical usability. For example, controlling for potentially endogenous variables, such as therapy utilization and functional status coded by HH providers, could increase R-squared but undermine the intent of the measure by masking variation it is intended to capture.
- Model discrimination and calibration results demonstrate good predictive ability across the full range of episodes, from low to high spending risk. There was no evidence of excessive under- or over-estimation at the extremes of episode risk.

And while there is no conceptual reason to target a certain level of R-squared, as noted above, the R-squared statistics are in line with other measures recently endorsed by NQF (e.g., #3510) and NQF does not set scientific acceptability standards based on R-squared minimum thresholds.

#3574 MSPB Clinician

We request that the Committee reconsider its overall 'do not recommend' evaluation of the MSPB Clinician measure as the must-pass validity criterion was not assessed in accordance with the correct NQF standards.

Reliability

We refer the Committee to the discussion under #3564 MSPB-PAC HH. We believe that the same considerations apply here for revisiting the CNR result of the reliability vote.

Validity

We request that the Committee reconsider the 'low' rating for validity as sub-criterion 2b was not correctly applied. First, face validity is allowable for new measure submissions, and second, we do not believe that the reasons given for the Committee's empirical validity concerns are sufficient to establish that the measure is unable to capture what it is intended to capture.

Face Validity

Sub-criterion 2b was not applied in accordance with the NQF's evaluation standards regarding face validity. If face validity is established, this should result in a 'moderate' validity rating:

- The NQF's evaluation standards in written materials state that face validity is an acceptable way to establish validity for new measure submissions.[11]
- Our measure submission details all the components of face validity in the evaluation algorithm:
    - The systematic assessment of validity involved recognized experts from a technical expert panel (TEP).
    - The TEP evaluated the extent to which the measure, as specified, can distinguish good and poor performance at the performance score-level through a 6-point Likert scale.
    - There was a high degree of consensus and substantial agreement amongst the TEP. We provided details of additional comments from the TEP, which indicated that there were no potential threats to validity.
- Members of the Committee indicated that they did not have concerns about the face validity of this measure.

Empirical Validity

Further to the above, we request that NQF reconsider its findings for empirical validity as it is unclear how sub-criteria 2b1 through 2b6 were applied to the measure so as to result in a 'low' rating. The fundamental question for assessing empirical validity is to assess whether the measure is able to correctly capture the intended measure concept; from our review of the meeting summary and Draft Report, we are unable to determine what aspects of the measure or empirical testing can substantiate this finding. For each of the four points raised in the Draft Report[12], we provide a response:

- *Attribution to multiple clinicians.* The discussion on this topic was primarily to clarify how the attribution rules work for medical and surgical MS-DRGs. We confirmed that episodes can be attributed to more than one TIN or TIN-NPI precisely to address the concern about singling out one provider to be responsible for an episode. We believe that the attribution rules do not constitute a threat to validity as the measure intent is to: (i) assess the costs of inpatient care at the clinician-level, and (ii) meet statutory requirements to have measures in the Merit-based Incentive Payment System (MIPS) cost performance category.
- *Specificity of episode window for different conditions.* While the episode window is fixed at 3-days pre-trigger and 30-days post trigger, the measure accounts for differences in types of inpatient care by: (i) risk-adjusting for MS-DRGs within each Major Diagnostic Category (MDC) to account for differences across hospitalizations, (ii) providing clinicians with a consistent post-discharge observation period, and (iii) aligning with quality measures that use a 30-day post-discharge period (e.g., MIPS 458 All-Cause Readmission, which captures all unplanned readmissions within 30 days of discharge). As the measure intent is to broadly capture inpatient care, we believe that using a consistent 30-day post-discharge period across hospitalizations correctly captures the intended concept.

---

[11] NQF, Committee Guidebook for the NQF Measure Endorsement Process v6.0 (Sept 2019), pages 44, 52-53. NQF, Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement (Sept 2019), page 38, NQF 'What good looks like', pages 9-11

[12] NQF, Cost and Efficiency, Spring 2020 Cycle: CDP Report. Draft Report for Comment (Aug 2020), pages 13, 41-42

- *Strength of correlations between predicted and risk-adjusted costs with types of services*. Our understanding is that the Committee's concern with 'low' correlations between some types of services ('clinical themes') and the risk-adjusted and predicted costs is that these results indicate that the model is not performing well. We redirect the Committee to the purpose of this analysis, which is to test whether the measure is responding to relative increases in higher cost services, not an overall assessment of the risk adjustment model's performance. Additionally, assuming high positive correlation for every clinical theme ignores differences in clinical context, cost efficiency, and care choices highlighted by the stratification. For example, increased use of preventative services or lower cost PAC settings can result in overall less cost downstream (i.e., we would expect to see an inverse relationship between risk-adjusted cost and those services' cost). For more costly settings (i.e., hospital readmissions and SNF), these results indicated that the measure is able to respond appropriately.
- *Performance of risk adjustment model and lack of SRF variables.* Please see discussion under the following measures as the same considerations apply here: #3564 MSPB-PAC HH: interpretation of R-squared in evaluating the performance of a risk adjustment model, and #3561 MSPB-PAC IRF: use of SRF in risk adjustment which – based on the inconsistent direction of the effect of SRF – can exacerbate bias, rather than reduce it.

#3575 TPCC

We request that the Committee consider the information summarized in the MSPB Clinician section in re-voting on the measure, as the must-pass validity criterion was not assessed in accordance with the correct NQF standards.

Face Validity

We refer the Committee to the discussion under #3574 MSPB Clinician regarding face validity. The same process was followed for the TPCC measure; as such, applying the validity evaluation criteria to this measure should lead to a 'moderate' rating.

Conclusion

CMS, with Acumen and Abt Associates, appreciate the opportunity to submit this request for the Committee to reconsider their evaluation of the Spring 2020 cycle cost measures in accordance with NQF guidance. We look forward to providing any further clarification needed at the Post-Comment Evaluation Meeting on October 1, 2020.

Appendix C: Developer Responses to Comments for Measures 3574 & 3575

NQF Measure 3574 Developer Responses

Thank you for your comment. Our responses below are structured to combine comments on similar topics. As the commenter noted, many of these were raised by other commenters and addressed during the Standing Committee evaluation meeting, so we have mainly focused on new points and briefly recapped where issues have already been discussed.

• Measure Intent: Scope and Attribution Options: The MSPB Clinician measure is intended to be a broad measure for overall cost of care surrounding an inpatient stay, as this is a significant contributor to overall Medicare spending. In this way, this population-based measure complements episode-based cost measures that are also used in MIPS, and is in line with other NQF endorsed measures that assess providers on all costs occurring during a particular timeframe. Importantly, this measure is used in MIPS alongside quality measures to capture other aspects of care, such as patient functional status, appropriate use of services, and mortality, amongst others.

As discussed during the Standing Committee evaluation meeting, this measure is used in MIPS which as a program offers options for participation at the individual or group level. Clinicians can choose how to participate based on their circumstances and preferences.

• Construction Logic: Capturing Cost: To clarify, the TPCC and MSPB Clinician measures use payment-standardized costs. This means that claim Medicare allowed amounts used in measure calculation are standardized to remove adjustments made differences in regional labor, geographic practice cost indices, and payments that support larger Medicare programs for any particular service. The result maintains differences that exist in actual payment resulting from choice of which services are provided.

• Applicable Measure Reliability Standards: This was discussed as an overarching issue during the Standing Committee evaluation meeting across many of the cost measures, and is covered in more detail in our Request for Reconsideration, attached as an appendix. There is no clear cut standard for reliability from the literature summarized in our submission materials or set by NQF who repeatedly declines to set one; CMS however has established 0.4 as the minimum standard in many quality reporting and value-based purchasing programs to strike an appropriate balance between reliability and measuring providers. As noted during the evaluation meeting, in the absence of definitive alternate standards and the importance of consistency, we point to numerous NQF endorsed measures with mean reliability values substantially below the reliability at the 25th percentile for the MSPB Clinician measure, including as recently as measures passed by CSAC in July 2020 (see for example #0684, #0686, #0073, #0689, #3032, #2561, #0694, #0669, and #0425).

The Standing Committee noted during the evaluation meeting that reliability depends on context and use. An important point to note is that these concerns are based on values at or below the 25th percentile for a subset of TINs with a single reporting clinician, representing approximately 7.5% of the total population. At the 25th percentile for all TIN-NPIs reliability was 0.60, and for TINs with one TIN-NPI the 25th percentile for reliability was 0.62. These values are above the mean reliability for

other endorsed measures. Overall, our testing demonstrated that there was good or high measure score reliability overall (TIN: 0.78 and TIN-NPI: 0.70). While raising the case minimum would increase the reliability, it is important to note that this comes at the expense of reducing the number of providers who are covered by the measure. The determination of a case minimum is a reporting decision for each measure's use in a program, so can change if the evidence supports such a decision.

• Measures Providing Meaningful Information: The intent of cost measures is to provide increased transparency to clinicians about their cost performance and drive clinicians to make more informed decisions about the costs and benefits of the services that they provide. As discussed during the evaluation meeting, CMS has been taking into account stakeholder feedback about what information is most useful for clinicians to be able to make informed care choices. For example, the MIPS 2018 performance period reports gave clinicians beneficiary and episode-level information, with detailed breakdowns of cost to understand what is driving costs for each case. This measure is not currently publicly reported, but any decisions to report this publicly will consider the points raised by the commenter to ensure that it is comprehensible and meaningful to patients.

• Accounting for provider characteristics: Rural Care: The risk adjustment model includes a robust set of clinical and other characteristics. Our testing indicates that the performance of urban and rural providers is similar across the distribution of measure scores, showing that rural providers are not being systematically disadvantaged by factors specific to rural care, such as more limited referral choices.

• Correlation with Quality Measures: We describe our approach to identifying quality measures with a conceptual relationship with the cost measure in section 1b1.2 of the testing form. There must be a conceptual basis for an expected relationship, including consideration of:
o Submission method
o Attributed clinician
o Setting
o Dimension of care being assessed (e.g., process vs outcome measure)
o Availability of data due to voluntary quality measure reporting

While we were limited due to small numbers of providers with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for MSPB Clinician and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.180. These results indicate that better performance on MSPB Clinician (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations.

• Concern about episodes being attributed to multiple clinicians across measures: A patient can be attributed to multiple providers within and across the MSPB Clinician, TPCC, and episode-based cost measures to ensure cost effectiveness of all individual clinicians or clinician groups managing the patient's care. It is also important to note that the scope of the MSPB Clinician, TPCC, and episode-based cost measures differs. The MSPB Clinician measure looks at the cost performance of clinicians providing care at inpatient hospitals, while TPCC focuses on primary care management outside the

inpatient setting. Meanwhile, episode-based measures only include costs related to the episode for a clinical condition or procedure and are focused on the clinician's specific role, as opposed to including all services that are provided to a patient over a given timeframe. Within a single cost measure, multiple clinicians involved in a single patient's care are each measured individually to ensure joint accountability of the patient's management. The measure calculation risk adjusts each clinician's observed costs for the patient with the same observable characteristics among their peers, rather than to a pre-defined standard. By comparing clinicians to their peers, who are all attributed in the same way, and measuring all clinicians who are responsible for the patient's care, we can expect this comparison to be fair. Across different cost measures, each measure will assess the specific role of the clinician care for the defined scope. As patients receive care across these different scopes, claims relevant to each measure will again be used in both measurements to accurately characterize the care for the individual cost measure. The measures are calculated separately, and then averaged into a single score for the MIPS Cost performance category. This approach ensures clinicians will not be double-penalized or rewarded for a high or low cost patient. In the aggregation of a MIPS cost performance score, the relative impact of a high or low cost patient in each cost measure is averaged for a given clinician or clinician group, rather than simply counted twice. This avoids compounding good or poor results, and allows the measure to accurately reflect clinician performance within the context of each individual measure. • Exclusions: Episodes Ending in Death. The measure excludes these episodes as they may not accurately reflect a clinician's performance. They could be unusually high-cost, due to perimortem treatment costs, or unusually low-cost, due to the truncated episode window. Neither of these cases accurately reflects the efficiency of the clinician performing the treatment. Additionally, it is important to keep in mind that the Cost performance category is one of four performance categories in MIPS, so performance in the other categories, including quality, will play an important role in assessing clinicians' overall performance and will ensure that poor care is penalized.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the scientific acceptability of the measure. We agree that the use of the measure in MIPS should allow providers to understand and make distinctions in care. The intent of cost measures is to provide increased transparency to clinicians about their cost performance and drive clinicians to make more informed decisions about the costs and benefits of the services that they provide. As discussed during the evaluation meeting, CMS has been taking into account stakeholder feedback about what information is most useful for clinicians to be able to make informed care choices. For example, the MIPS 2018 performance period reports gave clinicians beneficiary and episode-level information, with detailed breakdowns of cost to understand what is driving costs for each case.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the reliability and validity of the measure. • Measure Intent: Scope: The MSPB Clinician measure is intended to be a broad measure for overall

cost of care surrounding an inpatient stay, as this is a significant contributor to overall Medicare spending. In this way, this population-based measure complements episode-based cost measures that are also used in MIPS, and is line with other NQF endorsed measures that assess providers on all costs occurring during a particular timeframe. Importantly, this measure is used in MIPS alongside quality measures to capture other aspects of care, such as patient functional status, appropriate use of services, and mortality, amongst others. We also clarify that the MSPB Clinician measure has undergone comprehensive re-evaluation specifically for use in MIPS, similar to the process used for the episode-based cost measures that the commenter mentions. This has included working with a Technical Expert Panel (TEP) and field testing this measure on a national scale.

• Concern about episodes being attributed to multiple clinicians across measures: A patient can be attributed to multiple providers within and across the MSPB Clinician, TPCC, and episode-based cost measures to ensure cost effectiveness of all individual clinicians or clinician groups managing the patient's care. It is also important to note that the scope of the MSPB Clinician, TPCC, and episode-based cost measures differs. The MSPB Clinician measure looks at the cost performance of clinicians providing care at inpatient hospitals, while TPCC focuses on primary care management outside the inpatient setting. Meanwhile, episode-based measures only include costs related to the episode for a clinical condition or procedure and are focused on the clinician's specific role, as opposed to including all services that are provided to a patient over a given timeframe.

Within a single cost measure, multiple clinicians involved in a single patient's care are each measured individually to ensure joint accountability of the patient's management. The measure calculation risk adjusts each clinician's observed costs for the patient with the same observable characteristics among their peers, rather than to a pre-defined standard. By comparing clinicians to their peers, who are all attributed in the same way, and measuring all clinicians who are responsible for the patient's care, we can expect this comparison to be fair.

Across different cost measures, each measure will assess the specific role of the clinician care for the defined scope. As patients receive care across these different scopes, claims relevant to each measure will again be used in both measurements to accurately characterize the care for the individual cost measure. The measures are calculated separately, and then averaged into a single score for the MIPS Cost performance category. This approach ensures clinicians will not be double-penalized or rewarded for a high or low cost patient. In the aggregation of a MIPS cost performance score, the relative impact of a high or low cost patient in each cost measure is averaged for a given clinician or clinician group, rather than simply counted twice. This avoids compounding good or poor results, and allows the measure to accurately reflect clinician performance within the context of each individual measure.

• Measures Providing Meaningful Information: The intent of cost measures is to provide increased transparency to clinicians about their cost performance and drive clinicians to make more informed decisions about the costs and benefits of the services that they provide. As discussed during the evaluation meeting, CMS has been taking into account stakeholder feedback about what information is most useful for clinicians to be able to make informed care choices. For example, the MIPS 2018

performance period reports gave clinicians beneficiary and episode-level information, with detailed breakdowns of cost to understand what is driving costs for each case.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the reliability, face validity, and empirical validity of the measure.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the reliability and validity of the measure. We did explore the correlation with quality measures, particularly outcomes. We describe our overall approach in section 1b1.2 of the testing form. There must be a conceptual basis for an expected relationship which can depend on many factors, including: o Submission method o Attributed clinician o Setting o Dimension of care being assessed (e.g., process vs outcome measure) o Availability of data due to voluntary quality measure reporting While we were limited due to small numbers of providers with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for MSPB Clinician and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.180. These results indicate that better performance on MSPB Clinician (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the reliability and validity of the measure. We clarify that the intent of the measure is to be a broad measure for overall cost of care surrounding an inpatient stay, as this is a significant contributor to overall Medicare spending. In this way, this population-based measure complements episode-based cost measures that are also used in MIPS, and is line with other NQF endorsed measures that assess providers on all costs occurring during a particular timeframe. Importantly, this measure is used in MIPS alongside quality measures to capture other aspects of care. Further, the intent of cost measures is to provide increased transparency to clinicians about their cost performance and drive clinicians to make more informed decisions about the costs and benefits of the services that they provide. As discussed during the evaluation meeting, CMS has been taking into account stakeholder feedback about what information is most useful for clinicians to be able to make informed care choices. For example, the MIPS 2018 performance period reports gave clinicians beneficiary and episode-level information, with detailed breakdowns of cost to understand what is

driving costs for each case. This measure is not currently publicly reported, but any decisions to report this publicly will consider the points raised by the commenter to ensure that it is comprehensible and meaningful to patients.

We appreciate the suggestions for additional service exclusions. To clarify, the service exclusions are done at the MDC level, so are intended to be quite limited in line with the measure intent as an assessment of overall costs of care and to ensure that exclusions do apply across all MS-DRGs within that MDC. For the examples that the commenter raised, there are cases where the inpatient admission is either due to radiation therapy, or radiation therapy is involved in treatment, so are not excluded across the MDC. We will however take these suggestions into account as part of ongoing measure maintenance for potential refinements.

NQF Measure 3575 Developer Responses

Thank you for your comment. Our responses below are structured to combine comments on similar topics. As the commenter noted, many of these were raised by other commenters and addressed during the Standing Committee evaluation meeting, so we have mainly focused on new points and briefly recapped where issues have already been discussed.

• Measure Intent: Scope and Attribution Options: The TPCC measure is intended to broadly measure total cost of care for patients as managed by their primary care provider. Primary care clinicians play an important role in managing the overall health of a patient, with both direct and indirect influence on total Medicare spending. In this way, this population-based measure complements episode-based cost measures that are also used in MIPS, and is line with other NQF endorsed measures that assess providers on all costs occurring during a particular timeframe. Importantly, this measure is used in MIPS alongside quality measures to capture other aspects of care, such as patient functional status, appropriate use of services, and mortality, amongst others.

As discussed during the Standing Committee evaluation meeting, this measure is used in MIPS which as a program offers options for participation at the individual or group level. Clinicians can choose how to participate based on their circumstances and preferences

• Construction Logic: Capturing Cost: To clarify, the TPCC and MSPB Clinician measures use payment-standardized costs. This means that claim Medicare allowed amounts used in measure calculation are standardized to remove adjustments made differences in regional labor, geographic practice cost indices, and payments that support larger Medicare programs for any particular service. The result maintains differences that exist in actual payment resulting from choice of which services are provided.

• Applicable Measure Reliability Standards: The Scientific Methods Panel, whose role is to provide consistency and expertise in the scientific acceptability of measures, passed the measure on this criterion. The Standing Committee also passed the measure on this criterion and raised no substantive issues on reliability in either the measure evaluation meeting or the draft report.

This was discussed as an overarching issue during the Standing Committee evaluation meeting across many of the cost measures, and is covered in more detail in our Request for Reconsideration, attached as an appendix. There is no clear cut standard for reliability from the literature summarized in our submission materials or set by NQF who repeatedly declines to set one; CMS however has established 0.4 as the minimum standard in many quality reporting and value-based purchasing programs to strike an appropriate balance between reliability and measuring providers. As noted during the evaluation meeting, in the absence of definitive alternate standards and the importance of consistency, we point to numerous NQF endorsed measures with mean reliability values substantially below the reliability at the 25th percentile for the TPCC measure, including as recently as measures passed by CSAC in July 2020 (see for example #0684, #0686, #0073, #0689, #3032, #2561, #0694, #0669, and #0425).

The Standing Committee noted during the evaluation meeting that reliability depends on context and use. An important point to note is that these concerns are based on values at or below the 25th percentile for a subset of TINs with a single reporting clinician, representing approximately 12.7% of the total population. At the 25th percentile for all TIN-NPIs reliability was 0.83, and for TINs with one TIN-NPI the 25th percentile for reliability was 0.73. These values are above the mean reliability for other endorsed measures. Overall, our testing demonstrated that there was good or high measure score reliability overall (TIN: 0.84 and TIN-NPI: 0.88). While raising the case minimum would increase the reliability, it is important to note that this comes at the expense of reducing the number of providers who are covered by the measure. The determination of a case minimum is a reporting decision for each measure's use in a program, so can change if the evidence supports such a decision.

• Measures Providing Meaningful Information: The intent of cost measures is to provide increased transparency to clinicians about their cost performance and drive clinicians to make more informed decisions about the costs and benefits of the services that they provide. As discussed during the evaluation meeting, CMS has been taking into account stakeholder feedback about what information is most useful for clinicians to be able to make informed care choices. For example, the MIPS 2018 performance period reports gave clinicians beneficiary level information, with detailed breakdowns of cost to understand what is driving costs for each case. This measure is not currently publicly reported, but any decisions to report this publicly will consider the points raised by the commenter to ensure that it is comprehensible and meaningful to patients.

• Accounting for Provider Characteristics: Rural Care: The risk adjustment model includes a robust set of clinical and other characteristics. Our testing indicates that the performance of urban and rural providers is similar across the distribution of measure scores, showing that rural providers are not being systematically disadvantaged by factors specific to rural care, such as more limited referral choices.

• Correlation with Quality Measures: We describe our approach to identifying quality measures with a conceptual relationship with the cost measure in section 1b1.2 of the testing form. In order to explore the relationship between a cost and quality measure, there must be a conceptual basis for an expected relationship. This relationship can depend on many factors, including:
o Submission method, as we prioritized non-claims based measures

o Attributed clinician
o Setting
o Dimension of care being assessed (e.g., process vs outcome measure)
o Availability of data due to voluntary quality measure reporting

For TPCC, we examined the potential for analyzing a relationship between TPCC and potential claims and non-claims based outcome measures. While we were limited due to small numbers of providers with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for TPCC and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.056. These results indicate that better performance on TPCC (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations.

• Concern about beneficiaries being attributed to multiple clinicians across measures: A patient can be attributed to multiple providers within and across the MSPB Clinician, TPCC, and episode-based cost measures to ensure cost effectiveness of all individual clinicians or clinician groups managing the patient's care. It is also important to note that the scope of the MSPB Clinician, TPCC, and episode-based cost measures differs. The MSPB Clinician measure looks at the cost performance of clinicians providing care at inpatient hospitals, while TPCC focuses on primary care management outside the inpatient setting. Meanwhile, episode-based measures only include costs related to the episode for a clinical condition or procedure and are focused on the clinician's specific role, as opposed to including all services that are provided to a patient over a given timeframe.

Within a single cost measure, multiple clinicians involved in a single patient's care are each measured individually to ensure joint accountability of the patient's management. The measure calculation risk adjusts each clinician's observed costs for the patient with the same observable characteristics among their peers, rather than to a pre-defined standard. By comparing clinicians to their peers, who are all attributed in the same way, and measuring all clinicians who are responsible for the patient's care, we can expect this comparison to be fair.

Across different cost measures, each measure will assess the specific role of the clinician care for the defined scope. As patients receive care across these different scopes, claims relevant to each measure will again be used in both measurements to accurately characterize the care for the individual cost measure. The measures are calculated separately, and then averaged into a single score for the MIPS Cost performance category. This approach ensures clinicians will not be double-penalized or rewarded for a high or low cost patient. In the aggregation of a MIPS cost performance score, the relative impact of a high or low cost patient in each cost measure is averaged for a given clinician or clinician group, rather than simply counted twice. This avoids compounding good or poor results, and allows the measure to accurately reflect clinician performance within the context of each individual measure.

• Episodes Ending in Death. For the TPCC measure, we do not exclude beneficiary months for beneficiaries who died during the performance period.

Thank you for your comment. Our responses below are structured to combine comments on similar topics. As the commenter noted, many of these were raised by other commenters and addressed during the Standing Committee evaluation meeting, so we have mainly focused on new points and briefly recapped where issues have already been discussed.

• Applicable Measure Reliability Standards: The Scientific Methods Panel, whose role is to provide consistency and expertise in the scientific acceptability of measures, passed the measure on this criterion. The Standing Committee also passed the measure on this criterion and raised no substantive issues on reliability in either the measure evaluation meeting or the draft report.

This was discussed as an overarching issue during the Standing Committee evaluation meeting across many of the cost measures, and is covered in more detail in our Request for Reconsideration, attached as an appendix. There is no clear cut standard for reliability from the literature summarized in our submission materials or set by NQF who repeatedly declines to set one; CMS however has established 0.4 as the minimum standard in many quality reporting and value-based purchasing programs to strike an appropriate balance between reliability and measuring providers. As noted during the evaluation meeting, in the absence of definitive alternate standards and the importance of consistency, we point to numerous NQF endorsed measures with mean reliability values substantially below the reliability at the 25th percentile for the TPCC measure, including as recently as measures passed by CSAC in July 2020 (see for example #0684, #0686, #0073, #0689, #3032, #2561, #0694, #0669, and #0425).

The Standing Committee noted during the evaluation meeting that reliability depends on context and use. An important point to note is that these concerns are based on values at or below the 25th percentile for a subset of TINs with a single reporting clinician, representing approximately 12.7% of the total population. At the 25th percentile for all TIN-NPIs reliability was 0.83, and for TINs with one TIN-NPI the 25th percentile for reliability was 0.73. These values are above the mean reliability for other endorsed measures. Overall, our testing demonstrated that there was good or high measure score reliability overall (TIN: 0.84 and TIN-NPI: 0.88). While raising the case minimum would increase the reliability, it is important to note that this comes at the expense of reducing the number of providers who are covered by the measure. The determination of a case minimum is a reporting decision for each measure's use in a program, so can change if the evidence supports such a decision.

• Correlation with Quality Measures: We describe our approach to identifying quality measures with a conceptual relationship with the cost measure in section 1b1.2 of the testing form. In order to explore the relationship between a cost and quality measure, there must be a conceptual basis for an expected relationship. This relationship can depend on many factors, including:
o Submission method, as we prioritized non-claims based measures
o Attributed clinician
o Setting
o Dimension of care being assessed (e.g., process vs outcome measure)
o Availability of data due to voluntary quality measure reporting

For TPCC, we examined the potential for analyzing a relationship between TPCC and potential claims and non-claims based outcome measures. While we were limited due to small numbers of providers

with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for MSPB Clinician and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.056. These results indicate that better performance on TPCC (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations.

• Attributing Beneficiaries: We appreciate the examples provided by ACG, but believe that this attribution method is effective at identifying primary care relationships between patients and clinicians. First, the triggering methodology is effective at identifying a primary care relationship since it requires two claims. Requiring multiple claims within a defined, relatively short time period avoids attribution from a single claim and ensures evidence of a sustained relationship, using multiple codes indicative of overall health care evaluation and management. Second, specialty and service category exclusions are applied to further protect against misattribution. Last, Table 6 of the testing form shows that attributed TINs and TIN-NPIs bill a substantial proportion of patient E&M claims related to primary care, as was intended. This indicates a strong relationship between attributed TINs and TIN-NPIs and the beneficiaries they treat.

• Measure Intent: Scope and Attribution Options: The TPCC measure is intended to broadly measure total cost of care for patients as managed by their primary care provider. Primary care clinicians play an important role in managing the overall health of a patient, with both direct and indirect influence on total Medicare spending. In this way, this population-based measure complements episode-based cost measures that are also used in MIPS, and is in line with other NQF endorsed measures that assess providers on all costs occurring during a particular timeframe. Importantly, this measure is used in MIPS alongside quality measures to capture other aspects of care, such as patient functional status, appropriate use of services, and mortality, amongst others.

As discussed during the Standing Committee evaluation meeting, this measure is used in MIPS which as a program offers options for participation at the individual or group level. Clinicians can choose how to participate based on their circumstances and preferences.

Multiple clinicians involved in a single patient's care are each measured individually to ensure joint accountability of the patient's management. The measure calculation risk adjusts each clinician's observed costs for the patient with the same observable characteristics among their peers, rather than to a pre-defined standard. By comparing clinicians to their peers, who are all attributed in the same way, and measuring all clinicians who are responsible for the patient's care, we can expect this comparison to be fair.

Holding multiple clinician groups that demonstrate responsibility for the patient accountable encourages coordination of care, communication, and multidisciplinary approaches to providing high quality care to the patient. Overall, both patients and clinicians benefit when all providers involved in the care of the patient are covered by similar incentives.

Thank you for your comment. Our responses below are structured to combine comments on similar topics. As the commenter noted, many of these were raised by other commenters and addressed during the Standing Committee evaluation meeting, so we have mainly focused on new points and briefly recapped where issues have already been discussed.

• Applicable Measure Reliability Standards: These Scientific Methods Panel, whose role is to provide consistency and expertise in the scientific acceptability of measures, passed the measure on this criterion. The Standing Committee also passed the measure on this criterion and raised no substantive issues on reliability in either the measure evaluation meeting or the draft report.

This was discussed as an overarching issue during the Standing Committee evaluation meeting across many of the cost measures, and is covered in more detail in our Request for Reconsideration, attached as an appendix. There is no clear cut standard for reliability from the literature summarized in our submission materials or set by NQF who repeatedly declines to set one; CMS however has established 0.4 as the minimum standard in many quality reporting and value-based purchasing programs to strike an appropriate balance between reliability and measuring providers. As noted during the evaluation meeting, in the absence of definitive alternate standards and the importance of consistency, we point to numerous NQF endorsed measures with mean reliability values substantially below the reliability at the 25th percentile for the TPCC measure, including as recently as measures passed by CSAC in July 2020 (see for example #0684, #0686, #0073, #0689, #3032, #2561, #0694, #0669, and #0425).

The Standing Committee noted during the evaluation meeting that reliability depends on context and use. An important point to note is that these concerns are based on values at or below the 25th percentile for a subset of TINs with a single reporting clinician, representing approximately 12.7% of the total population. At the 25th percentile for all TIN-NPIs reliability was 0.83, and for TINs with one TIN-NPI the 25th percentile for reliability was 0.73. These values are above the mean reliability for other endorsed measures. Overall, our testing demonstrated that there was good or high measure score reliability overall (TIN: 0.84 and TIN-NPI: 0.88). While raising the case minimum would increase the reliability, it is important to note that this comes at the expense of reducing the number of providers who are covered by the measure. The determination of a case minimum is a reporting decision for each measure's use in a program, so can change if the evidence supports such a decision.

• Correlation with Quality Measures: We describe our approach to identifying quality measures with a conceptual relationship with the cost measure in section 1b1.2 of the testing form. In order to explore the relationship between a cost and quality measure, there must be a conceptual basis for an expected relationship. This relationship can depend on many factors, including:
o Submission method, as we prioritized non-claims based measures
o Attributed clinician
o Setting
o Dimension of care being assessed (e.g., process vs outcome measure)
o Availability of data due to voluntary quality measure reporting

For TPCC, we examined the potential for analyzing a relationship between TPCC and potential claims and non-claims based outcome measures. While we were limited due to small numbers of providers

with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for TPCC and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.056. These results indicate that better performance on TPCC (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations. • Attributing Beneficiaries: We appreciate the examples provided by ACG, but believe that this attribution method is effective at identifying primary care relationships between patients and clinicians. First, the triggering methodology is effective at identifying a primary care relationship since it requires two claims. Requiring multiple claims within a defined, relatively short time period avoids attribution from a single claim and ensures evidence of a sustained relationship, using multiple codes indicative of overall health care evaluation and management. Second, specialty and service category exclusions are applied to further protect against misattribution. Last, Table 6 of the testing form shows that attributed TINs and TIN-NPIs bill a substantial proportion of patient E&M claims related to primary care, as was intended. This indicates a strong relationship between attributed TINs and TIN-NPIs and the beneficiaries they treat. • Oncologist Exclusions: While radiation oncologists are excluded from the TPCC measure, medical oncologists are included in the measure. This is because medical oncologists see patients throughout the duration of their cancer care, even after they have received treatment. Additionally, they are included because these providers see recovered patients and the presence of cancer is risk adjusted for in the risk adjustment model.	
Thank you for your comment. Our responses below are structured to combine comments on similar topics. As the commenter noted, many of these were raised by other commenters and addressed during the Standing Committee evaluation meeting, so we have mainly focused on new points and briefly recapped where issues have already been discussed. • Applicable Measure Reliability Standards: The Scientific Methods Panel, whose role is to provide consistency and expertise in the scientific acceptability of measures, passed the measure on this criterion. The Standing Committee also passed the measure on this criterion and raised no substantive issues on reliability in either the measure evaluation meeting or the draft report. This was discussed as an overarching issue during the Standing Committee evaluation meeting across many of the cost measures, and is covered in more detail in our Request for Reconsideration, attached as an appendix. There is no clear cut standard for reliability from the literature summarized in our submission materials or set by NQF who repeatedly declines to set one; CMS however has established 0.4 as the minimum standard in many quality reporting and value-based purchasing programs to strike an appropriate balance between reliability and measuring providers. As noted during the evaluation meeting, in the absence of definitive alternate standards and the importance of consistency, we point to numerous NQF endorsed measures with mean reliability values substantially below the reliability at the 25th percentile for the TPCC measure, including as recently as measures passed by CSAC in July 2020 (see for example #0684, #0686, #0073, #0689, #3032, #2561, #0694,	

#0669, and #0425).

The Standing Committee noted during the evaluation meeting that reliability depends on context and use. An important point to note is that these concerns are based on values at or below the 25th percentile for a subset of TINs with a single reporting clinician, representing approximately 12.7% of the total population. At the 25th percentile for all TIN-NPIs reliability was 0.83, and for TINs with one TIN-NPI the 25th percentile for reliability was 0.73. These values are above the mean reliability for other endorsed measures. Overall, our testing demonstrated that there was good or high measure score reliability overall (TIN: 0.84 and TIN-NPI: 0.88). While raising the case minimum would increase the reliability, it is important to note that this comes at the expense of reducing the number of providers who are covered by the measure. The determination of a case minimum is a reporting decision for each measure's use in a program, so can change if the evidence supports such a decision.

• Correlation with Quality Measures: We describe our approach to identifying quality measures with a conceptual relationship with the cost measure in section 1b1.2 of the testing form. In order to explore the relationship between a cost and quality measure, there must be a conceptual basis for an expected relationship. This relationship can depend on many factors, including:
o Submission method, as we prioritized non-claims based measures
o Attributed clinician
o Setting
o Dimension of care being assessed (e.g., process vs outcome measure)
o Availability of data due to voluntary quality measure reporting

For TPCC, we examined the potential for analyzing a relationship between TPCC and potential claims and non-claims based outcome measures. While we were limited due to small numbers of providers with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for TPCC and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.056. These results indicate that better performance on TPCC (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations.

• Risk Adjustment and Stinting of Care: We understand the concern that some clinicians or clinician groups might limit the number of services they render to patients to avoid getting attributed sicker patients. However, this potential concern is addressed by risk adjustment, which creates a level field for clinicians by accounting for patient complexity through examination of previous services and diagnoses that are predictive of high episode spending. As such, the measures are able to more accurately represent clinician performance across a broad patient case-mix, and ensure that there is no incentive to avoid providing care to complex patients. In addition, the potential concern of stinting of care in general is addressed primarily by including costs for services that occur as a consequence of care decisions, such as complications, in the measure calculation. Finally, the Cost performance category is one of four performance categories in MIPS, so performance in the other categories, including quality, will play an important role in assessing clinicians' overall performance.

• Transition of Care: The intent of the TPCC measure is to capture primary care relationships, which

by their nature, are long-term and have effects beyond when a patient has changed providers. For example, we would want to capture the costs of downstream services that are related to the scope of primary care (e.g., preventive care). By monitoring the total cost of care over a longer period of time (i.e., 12 months), the measure can provide a more complete picture of provider care. Additionally, holding multiple clinician groups that demonstrate responsibility for the patient accountable encourages coordination of care, communication, and multidisciplinary approaches to providing high quality care to the patient. Overall, both patients and clinicians benefit when all providers involved in the care of the patient are covered by similar incentives.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the reliability and validity of the measure. We did explore the correlation with quality measures, particularly outcomes. We describe our overall approach in section 1b1.2 of the testing form. There must be a conceptual basis for an expected relationship which can depend on many factors, including: o Submission method o Attributed clinician o Setting o Dimension of care being assessed (e.g., process vs outcome measure) o Availability of data due to voluntary quality measure reporting While we were limited due to small numbers of providers with both a cost measure and quality measure that has a conceptual relationship with the cost measure, we conducted a correlation for TPCC and MIPS #448 All-Cause Readmission: Unplanned Readmission within 30 Days of Discharge at the TIN-level. We expected fewer readmissions would correlate with lower cost as hospital admissions are costly. There is a statistically significant positive association with a correlation of 0.056. These results indicate that better performance on TPCC (i.e., lower cost) correlates with fewer unplanned readmissions, in line with expectations.
Thank you for your comment. We refer to our Request for Reconsideration included in the Appendix for our overall responses to the Committee's concerns on the reliability and validity of the measure. We clarify that the intent of the TPCC measure is intended to broadly measure total cost of care for patients as managed by their primary care provider. Primary care clinicians play an important role in managing the overall health of a patient, with both direct and indirect influence on total Medicare spending. In this way, this population-based measure complements episode-based cost measures that are also used in MIPS, and is line with other NQF endorsed measures that assess providers on all costs occurring during a particular timeframe. Importantly, this measure is used in MIPS alongside quality measures to capture other aspects of care. Further, the intent of cost measures is to provide increased transparency to clinicians about their cost performance and drive clinicians to make more informed decisions about the costs and benefits of the services that they provide. As discussed during the evaluation meeting, CMS has been taking into account stakeholder feedback about what information is most useful for clinicians to be able to make

informed care choices. For example, the MIPS 2018 performance period reports gave clinicians beneficiary level information, with detailed breakdowns of cost to understand what is driving costs for each case. This measure is not currently publicly reported, but any decisions to report this publicly will consider the points raised by the commenter to ensure that it is comprehensible and meaningful to patients.