Cost and Resource Use 2016-2017

FINAL TECHNICAL REPORT

August 30, 2017

This report is funded by the Department of Health and Human Services under contract HHSM-500-2012-00009I Task Order HHSM-500-T0008.

NATIONAL QUALITY FORUM

Contents

Cost and Resource Use 2016-2017

FINAL TECHNICAL REPORT

Executive Summary

In 2015, healthcare spending in the United States reached $3.2 trillion or approximately $9,990 per person.[1] This represented a 5.8 percent increase over 2014 spending levels.[2] Despite this high level of spending, the U.S. continues to rank below other developed countries for health outcomes including lower life expectancy and greater prevalence of chronic diseases.[3] Healthcare quality is also an issue with the U.S. falling behind other developed countries in the quality domains of effective care, safe care, coordinated care, and patient-centered care.[4] The factors contributing to these concerning trends are as complex as the healthcare system itself and include physician practice patterns, regional market influences, and access to care. Improving efficiency has the potential to simultaneously reduce the rate of cost growth and improve the quality of care provided.

As reducing costs continues to be a focus of healthcare reform, it is important to understand the current use of resources in the healthcare system as it relates to quality—especially how resource use relates to health outcomes. Recent legislation—including the Improving Medicare Post-Acute Care Transformation Act (IMPACT) of 2014, and the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)—requires the use of resource use measures to support payment reform efforts. The results of resource use measures will also be included on the physician compare website, and will ultimately be included in the Merit-Based Incentive Payment System (MIPS) for physicians. 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 value.

For this project, the Cost and Resource Use Standing Committee evaluated three measures undergoing maintenance review against NQF's evaluation criteria. The Standing Committee recommended all three measures, and the Consensus Standards Approval Committee (CSAC) ratified the continued endorsement of these measures:

- 1598 Total Resource Use Population-Based PMPM Index
- 1604 Total Cost of Care Population-Based PMPM Index
- 2158 Medicare Spending Per Beneficiary (MSPB) – Hospital

Brief summaries of the measures and Committee discussion are included in the body of the report; detailed summaries of the Committee's discussion, ratings of the criteria for each measure, and public and member comments are in Appendix A.

Introduction

In 2015, healthcare spending in the United States reached $3.2 trillion or approximately $9,990 per person.[5] This represented a 5.8 percent increase over 2014 spending levels, and expenditures related to private health insurance, hospital care, physician services, and clinical services were the primary contributors.[6] Despite this high level of spending, the U.S. continues to rank below other developed countries for health outcomes including lower life expectancy and greater prevalence of chronic diseases.[7] Healthcare quality is also an issue with the U.S. falling behind other developed countries in the quality domains of effective care, safe care, coordinated care, and patient-centered care.[8] The factors contributing to these concerning trends are as complex as the healthcare system itself and include physician practice patterns, regional market influences, and access to care. Improving efficiency has the potential to simultaneously reduce the rate of cost growth and improve the quality of care provided.

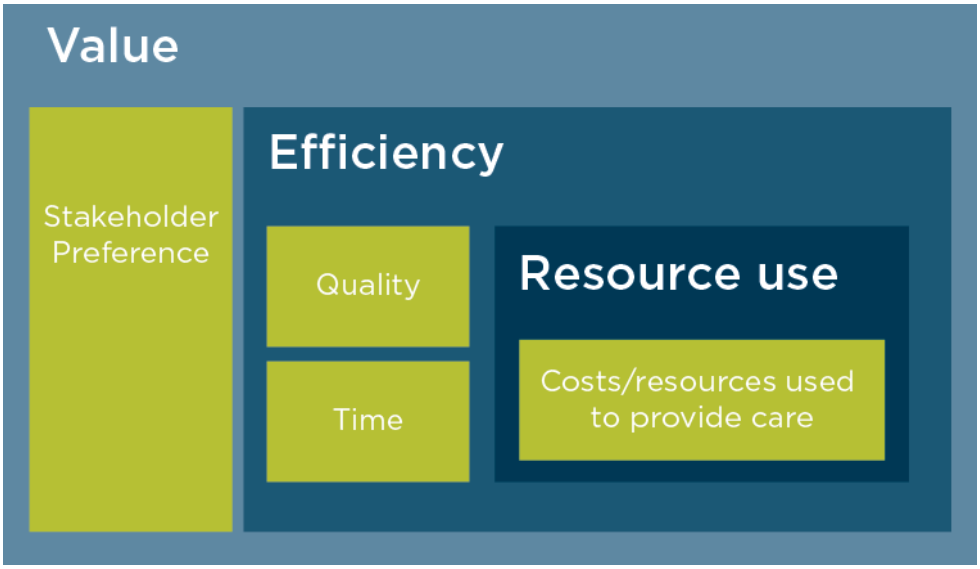
As reducing costs continues to be a focus of healthcare reform, it is important to understand the current use of resources in the healthcare system as it relates to quality—especially how resource use relates to health outcomes. Recent legislation—including the Improving Medicare Post-Acute Care Transformation Act (IMPACT) of 2014, and the Medicare Access & CHIP Reauthorization Act of 2015 (MACRA)—requires the use of resource use measures to support payment reform efforts. Resource use measures will also be included on the physician compare website, and will ultimately be included in the Merit-based Incentive Payment System (MIPS) for physicians. 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.

For nearly a decade, the National Quality Forum (NQF) has been working to advance cost and resource use measurement. In January 2010, NQF released a report, Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care, which addressed cost and resource use as one of the three overarching domains for assessing efficiency. NQF defined efficiency 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. The report's framework advised that measures of resource use and cost should acknowledge the value of measuring actual prices paid and standardized prices, in addition to measuring overall use.

Resource use measures can be defined as measures of health services applied to a population or event.[9] A resource use measure counts the frequency of use of defined health system resources, and some may further apply a dollar amount (e.g., allowable charges or standardized prices) to each unit of resource use. Alternatively, a cost of care measure calculates total healthcare spending. This includes the total resource use and the unit prices, by payer or consumer, for a healthcare service or group of services associated with a specified patient population, time period, and unit of clinical accountability.[10] Current approaches for measuring resource use and cost range from broadly focused measures, such as per capita measures, which address total healthcare spending or resource use per person, to those with a more narrow focus, such as measures dealing with the healthcare spending or resource use of an individual procedure (e.g., a hip replacement). This project builds on the 2010 measurement framework

and emphasizes that measures of cost, resource use, and quality must be aligned in order to truly understand efficiency and value (Figure 1).

Figure 1. Resource Use as a Building Block Toward Measuring Efficiency and Value.



This project represents the fourth phase of NQF's work on evaluating and endorsing cost and resource use measures. The prior three phases of work focused on the evaluation of both condition-specific and noncondition-specific measures of total cost, using both per capita or per hospitalization episode approaches. This fourth phase involved the review of three noncondition-specific measures of cost and resource use.

Trends and Performance

U.S. healthcare spending increased 5.8 percent in 2015 to reach 3.2 trillion dollars. The growth in spending was driven by coverage expansion as well as growth in spending for private health insurance, hospital care, physician and clinician services, Medicaid, and retail prescription drugs.[11] Hospital care accounted for the largest portion of expenditures at 1.0 trillion dollars or 32 percent. Physician and clinical services followed with 635 billion dollars of expenditures or 20 percent.[12]

NQF Portfolio of Performance Measures for Cost and Resource Use

The Cost and Resource Use Standing Committee (see Appendix C) oversees NQF's portfolio of six cost and resource use measures (see Table 1).

Table 1. NQF Cost and Resource Use Portfolio of Measures

NQF #	Title	Category
1598	Total Resource Use Population-Based PMPM Index	Noncondition-specific per capita resource use measure
1604	Total Cost of Care Population-Based PMPM Index	Noncondition-specific per capita cost measure
2431	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Acute Myocardial Infarction (AMI)	Condition-specific, episode-based cost measure
2436	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Heart Failure	Condition-specific, episode-based cost measure
2579	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Pneumonia	Condition-specific, episode-based cost measure
2158	Medicare Spending Per Beneficiary	Noncondition-specific, episode-based cost measure

National Quality Strategy

NQF-endorsed cost and resource use measures support the National Quality Strategy (NQS). The NQS serves as the overarching framework for guiding and aligning public and private efforts to improve the efficiency of healthcare in the U.S. The NQS establishes the "triple aim" of better care, affordable care, and healthy people/healthy communities. The NQF portfolio of cost and resource use measures specifically addresses the priority of making care more affordable for individuals, families, employers and government. These measures also support the development and spread of new healthcare delivery models.

Use of Measures in the Portfolio

NQF endorsement is valued because the evaluation is rigorous and transparent and conducted by multistakeholder committees. These committees are comprised of clinicians and other experts from the full range of healthcare providers, employers, health plans, public agencies, community coalitions, and patients—many of whom use measures on a daily basis to ensure better care. Moreover, NQF-endorsed measures undergo routine “maintenance” (i.e., re-evaluation) to ensure that they are still the best-available measures and reflect the current science. Importantly, federal law requires that preference be given to NQF-endorsed measures for use in several federal public reporting and performance-based payment programs.

Several of the measures in the portfolio are used in federal quality initiative programs, including the Hospital Inpatient Quality Reporting Program and Hospital Value-Based Purchasing Program. See Appendix B for further information on the use of these measures in federal programs.

Improving NQF's Cost and Resource Use Portfolio and Evaluation Process

Committee Input on Gaps in the Portfolio

During its discussions, the Committee identified areas for cost and resource use measure development, including:

- Total per capita cost measure for Medicare patients
- Measures for post-acute care settings, including home health, skilled nursing facilities, and long-term acute care.
- Measures that examine spending for high-cost, high-risk acute patients, including patients with multiple chronic diseases
- Measures that examine resource use across the patient episode of care – spanning across care settings, providers, and time

Committee Feedback on Cost and Resource Use Evaluation Criteria

The Cost and Resource Use Standing Committee discussed opportunities to update the Cost and Resource Use Evaluation Criteria in the future. These updates are intended to simplify the evaluation criteria and align with updates to the NQF Quality Measure Evaluation Criteria. A detailed description of the proposed changes appears in Appendix F. A summary of the proposed changes includes:

- Updating the Importance to Measure and Report criterion to clarify the following:
    - Remove specific language requiring candidate cost and resource measures to address a national health goal since cost and resource use measures typically address a critical element of the National Quality Strategy.
    - Clarify that the intent of the performance gap subcriterion is to examine the extent to which the measure helps identify disparities.
    - Remove language on evaluating the intent of the resource use measure since the Scientific Acceptability criterion includes how well the measure specifications align with the measure intent. Discussion of measure intent under Importance has often been redundant and unnecessary.
- Update the Scientific Acceptability criterion to align with the NQF Quality Measure Evaluation Criteria by adding requirements to submit ICD-10 and eMeasure specifications, when appropriate.

During the NQF member and public comment period, NQF received two public comments supporting the updates to the criteria and requesting more information on the updates to the performance gap criterion. The performance gap subcriterion is meant to address the question of whether a particular measure actually addresses a cost and resource use problem. Because the measurement enterprise is resource intensive, NQF's position is to endorse measures that address areas of known gaps in performance (i.e., those for which there is actually opportunity for improvement). Opportunity for improvement can be demonstrated by data that indicate overall poor performance (in the activity or outcome targeted by the measure), substantial variation in performance across providers, or variation in performance for certain subpopulations (i.e., disparities in care).

Cost and Resource Use Measure Evaluation

On March 15, 2017, the Cost and Resource Use Standing Committee evaluated three measures undergoing maintenance review against NQF's current cost and resource use measure evaluation criteria. To facilitate the evaluation, the Committee performed a preliminary review of the measures against the evaluation subcriteria via a preliminary evaluation survey, the results of which were included in the measure evaluation worksheets that were shared with the Committee and the public prior to the in-person meeting.

Table 2. Cost and Resource Use Measure Evaluation Summary

	Maintenance	New	Total
Measures under consideration	3	0	3
Measures endorsed	3	0	3

Comments Received Prior to Committee Evaluation

NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). In addition, NQF solicits comments prior to the evaluation of the measures via an online tool located on the project webpage. For this evaluation cycle, the pre-evaluation comment period was open from February 20 to March 6, 2017, for all three measures under review. The project received 33 pre-evaluation comments (Appendix E). Comments included questions about measure specifications, risk-adjustment methods, and interpretation of submitted performance data, as well as comments from healthcare organizations and practitioners expressing support for the re-endorsement of NQF #1598 and NQF #1604. NQF staff provided all submitted comments to the Committee prior to its initial deliberations during the in-person meeting.

Overarching Issues

During the discussion of the measures under review, the Committee raised two overarching issues that factored into the Committee's ratings and recommendations for the measures: risk adjustment for social risk factors and attribution.

Risk Adjustment for Social Risk Factors

Three major issues were identified related to social risk factors. First, the Committee examined the social risk factors tested in the risk-adjustment models of the candidate measures. The Committee discussed the need to test social risk factors capturing individual-level attributes and potentially community-level attributes. Some Committee members argued that patient-level risk factors should be favored in risk-adjustment models. However, others argued individual-level data may be difficult to capture and community-level factors should be explored when they can serve as an appropriate proxy. Committee members also highlighted the need to explore the impact of a person's community and the resources available and to consider adjusting for these factors. The Committee noted that when a person has fewer community resources available, the healthcare system may need to spend more to address his or her needs.

Second, the Committee discussed the impact of adjustment for social risk factors on different groups of providers. For each of the three measures under consideration, the inclusion of socioeconomic or sociodemographic variables did not result in statistically significant changes in measure scores for a high percentage of providers (90-97 percent). However, some Committee members urged developers to provide more information on those providers whose measure scores underwent a significant or larger than average change when sociodemographic or socioeconomic variables were included in the risk-adjustment models. Committee members stressed the need to examine and better understand the impact of risk adjustment on this subset of providers so that the implications of including or not including a given variable could be fully understood.

Finally, the Committee noted the need to better understand the role of unmeasured clinical complexity and how these factors may interact with a person's social risk factors. Committee members noted the unique nature of cost and resource use measures and the need to better understand how resources are used. Committee members recognized that those with social risk factors and those who are more medically complex may require more resources to achieve the same outcome as less vulnerable patients. The Committee stressed the importance of doing appropriate risk adjustment for cost and resource use measures to ensure that these measures do not worsen disparities, especially when they are used to determine payment through value-based purchasing.

Attribution

Cost and resource use measures are increasingly used in value-based purchasing programs. However, the use of these measures to reward or penalize providers requires an understanding of who is able to influence the costs of a person's care, as many parties are often involved in providing care. Attribution is the methodology used to assign patients, and their healthcare outcomes, to providers or clinicians. Appropriate attribution of a patient's healthcare costs has been an ongoing measurement challenge. Multiple clinicians and providers are frequently involved in a patient's care; however, a measure may assign responsibility for all of the costs for a certain time period or episode to one clinician or provider. For example, one measure reviewed by the Committee, NQF #2158, assesses the total spending per Medicare beneficiary immediately prior to, during, and following a patient's hospital stay. Some stakeholders have raised concerns about the measure's attribution strategy noting that the majority of variation in this measure is due to spending that occurs in the post-acute care settings. While the hospital has some influence, it may not have complete control over the source of variation in the measure.

The Committee reviewed guidance from NQF's recent attribution project and provided input on how the Expert Panel's guidance could be applied to cost and resource use measures. The Committee discussed the need for a measure's attribution guidelines to be clear and specific, but also flexible enough not to impede measure implementation. Given that patients may see multiple providers and provider types across care settings during their course of care, determining who is responsible for a given patient's resource use and health outcomes is difficult. During this round of measure evaluation, the Committee discussed the tension that exists between needing attribution guidelines to be both precise and flexible. Precision is needed to assist measure implementers in determining how to attribute patients consistently so as to allow for comparisons across providers, organizations, and over time.

The Committee recognized the important role measurement plays in understanding healthcare spending. The Committee cautioned that imperfect attribution should not impede progress towards better understanding healthcare costs. However, the Committee noted that the need to attribute costs must be balanced with the risk for unintended consequences. The Committee also noted the need for attribution models that support care coordination and team-based care as the system aims to transition from fee-for-service to population-based payment. The Committee suggested that attribution models better capture the role of nurse practitioners and physician assistants as a way to address the transition to team-based care.

Summary of Measure Evaluation

The following brief summaries of the measure evaluation highlight the major issues that the Committee considered. Details of the Committee's discussion, ratings of the criteria for each measure, and public and member comments are in Appendix A.

1598 Total Resource Use Population-Based PMPM Index (HealthPartners): Endorsed

Description: The Resource Use Index (RUI) is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services. A Resource Use Index when viewed together with the Total Cost of Care measure (NQF-endorsed #1604) provides a more complete picture of population based drivers of health care costs; **Measure Type**: Cost/Resource Use; **Level of Analysis**: Population: Community, County or City, Clinician: Group/Practice; **Setting of Care**: Hospital: Acute Care Facility, Ambulatory Surgery Center, Birthing Center, Clinician Office/Clinic, Hospital: Critical Care, Dialysis Facility, Emergency Department, Emergency Medical Services/Ambulance, Home Health, Hospice, Hospital, Imaging Facility, Behavioral Health: Inpatient, Inpatient Rehabilitation Facility, Laboratory, Long Term Acute Care, Nursing Home/SNF, Other, Behavioral Health: Outpatient, Outpatient Rehabilitation, Pharmacy, Urgent Care - Ambulatory; **Data Source**: Claims (Only)

Measure #1598 has been NQF-endorsed since January 2012. The only substantial change to this per capita measure of total resource use is increasing the truncation limit to $125,000. Truncation is a method used to limit costs above a certain value to reduce the impact of outliers. The Committee generally agreed that resource use continues to be an important area of measure focus with wide variation in performance. The Committee discussed how this measure could be used to drive quality improvement, and the developer clarified that the measure can be disaggregated to identify specific areas of opportunity. The Committee also encouraged the developer to consider expanding the measure to include the over age 65 Medicare population since that is an important cohort not included in the current measure.

The Committee reviewed updated reliability testing for the measure. Committee members raised some concerns around the attribution approach and the localized area used for testing. However, the Committee recognized that the measure is currently used widely and agreed it is reliable. Under the validity criterion, the Committee raised questions about the need to include social risk factors in the risk-adjustment model. The developer noted the limited impact of these factors on the performance of

the risk-adjustment model. Ultimately, the Committee agreed that the measure was valid but encouraged the developer to continue to explore the role of social risk.

The Committee agreed that the measure is feasible and usable, acknowledging the widespread use of this measure in public reporting programs, payment programs, and quality improvement programs. The Committee agreed that measure #1598 meets the NQF criteria and recommended it for continued endorsement. The CSAC ratified this endorsement.

1604 Total Cost of Care Population-Based PMPM Index (HealthPartners): Endorsed

Description: Total Cost of Care reflects a mix of complicated factors such as patient illness burden, service utilization and negotiated prices. Total Cost Index (TCI) is a measure of a primary care provider's risk adjusted cost effectiveness at managing the population they care for. TCI includes all costs associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services. A Total Cost Index when viewed together with the Total Resource Use measure (NQF-endorsed #1598) provides a more complete picture of population based drivers of health care costs. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Population: Community, County or City, Clinician: Group/Practice; **Setting of Care**: Hospital: Acute Care Facility, Ambulatory Surgery Center, Birthing Center, Clinician Office/Clinic, Hospital: Critical Care, Dialysis Facility, Emergency Department, Emergency Medical Services/Ambulance, Home Health, Hospice, Hospital, Imaging Facility, Behavioral Health: Inpatient, Inpatient Rehabilitation Facility, Laboratory, Long Term Acute Care, Nursing Home/SNF, Other, Behavioral Health: Outpatient, Outpatient Rehabilitation, Pharmacy, Urgent Care - Ambulatory **Data Source**: Claims (Only)

Measure #1604 has been NQF-endorsed since January 2012. Since its last endorsement, the measure's only substantial change is the truncation limit; the developer increased it from $100,000 to $125,000 in order to adjust for inflation and present medical costs. This per capita (population- or patient-based) measure calculates the total cost of care of a commercial population. When used alongside measure #1598, this measure provides information on population-based drivers of healthcare costs. The Committee agreed that this measure addresses an important aspect of healthcare.

The Committee agreed that the measure continues to demonstrate a high degree of reliability. It noted that the measure's construction and calculation logic and testing results remain strong. For this maintenance submission, the developer summarized updated validity testing conducted using provider data from 2014 and 2015. The Committee asked for clarification of how price is included and how different payment models are handled in the measure. The developer clarified that NQF #1604 is a total cost measure that includes the plan liability plus the member liability. The measure user can select the payment system (e.g., fee-for-service or DRG-based payment). Ultimately, the Committee agreed that the measure met the Scientific Acceptability criterion.

The Committee agreed that the measure is feasible and usable acknowledging the widespread use of this measure in public reporting programs, payment programs, and quality improvement programs. The Committee agreed that measure #1604 met the criteria and recommended it for continued endorsement. The CSAC ratified this endorsement.

2158 Medicare Spending Per Beneficiary (MSPB) - Hospital (Acumen, LLC/Centers for Medicare & Medicaid Services): Endorsed

Description: The Medicare Spending Per Beneficiary (MSPB) - Hospital measure evaluates hospitals' risk-adjusted episode costs relative to the risk-adjusted episode costs of the national median hospital. Specifically, the MSPB-Hospital measure assesses the cost to Medicare for services performed by hospitals and other healthcare providers during an MSPB-Hospital episode, which is comprised of the periods immediately prior to, during, and following a patient's hospital stay. The MSPB-Hospital measure is not condition specific and uses standardized prices when measuring costs. Beneficiary populations eligible for the MSPB-Hospital calculation include Medicare beneficiaries enrolled in Medicare Parts A and B who were discharged from short-term acute Inpatient Prospective Payment System (IPPS) hospitals during the period of performance. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Facility; **Setting of Care**: Hospital : Acute Care Facility; **Data Source**: Claims (Only), Other

Measure #2158 was first endorsed in December 2013. The Committee agreed that this measure addresses an important area of measurement given rising Medicare expenditures. While the Committee generally supported the reliability and validity of the measure, they did question the testing of the risk-adjustment model, specifically the need to include social risk factors. Committee members discussed the developer's choice to initially test only two sociodemographic variables—race (i.e., non-black and black) and income-to-poverty ratio at the five-digit zip code level. Committee members noted that race should not be used as a proxy for socioecomonic status per guidance from NQF's Disparities Standing Committee. Some Committee members stated that the income-to-poverty ratio at the five-digit zip code level is not precise enough to accurately represent an individual patient's socioeconomic status and therefore may not be appropriate for testing. In response to this concern, the developer presented information on additional risk-adjustment testing examining the effect of dual eligiblity for Medicare and Medicaid status on measure scores. Results from this additional testing indicated that the inclusion of the dual eligible status did not result in a significant change in measure scores for the majority of providers. The Committee strongly urged the developer to continue testing additional variables within the risk-adjustment approach.

The Committee agreed that the measure is feasible and widely used, but suggested that developers share more detailed information with providers in their measure summary reports (e.g., utilization rates by major diagnostic categories) as a way to help direct improvement efforts. However, the Committee noted that the measure is only endorsed at the facility level of analysis and cautioned that its use in clinician-level programs like the Merit-Based Incentive Payment System (MIPS) is not endorsed. The Committee encouraged CMS to test the measure at the clinician level of analysis and bring the additional testing back for the Committee's review in the future. Ultimately, the Committee agreed that the measure met the NQF criteria and recommended NQF #2158 for continued endorsement. The CSAC ratified the endorsement.

Comments Received After Committee Evaluation

After the Committee's evaluation of the three maintenance measures, NQF solicited comments on the draft report via an online tool from April 20, 2017, through May 19, 2017. During this period, NQF received 21 comments from nine member organizations. Four major themes were identified in the post-

evaluation comments about the measures, including concerns about (1) reliability and validity, (2) adjusting for social risk factors, (3) concerns about populations included in the measures, and (4) support for the measures. As noted above, commenters also expressed support for proposed updates to the cost and resource use measure evaluation criteria.

Risk Adjustment for Social Risk Factors

Four comments—two on measure #2158 *Medicare Spending Per Beneficiary (MSPB) – Hospital,* one on #1598 *Total Resource Use Population-Based PMPM Index*, and one on #1604 *Total Cost of Care Population-Based PMPM Index*—expressed concern regarding potentially insufficient adjustments made for social risk factors. Commenters were concerned that the developers did not provide an adequate conceptual basis and justification for the risk factors included in the testing, and did not include several factors commonly available in the literature. The comments submitted to NQF urged the Committee to take a more in-depth look at the need for SDS adjustment, given the potentially negative impact these measures could have on providers. Commenters encouraged additional testing of SDS factors.

The Committee agreed that consideration of social risk factors in risk-adjustment models is a critical issue in measurement science. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer. The Committee recognized that there continue to be limitations in the available data elements to capture unmeasured clinical and social risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between social risk factors and cost and resource use.

While the Committee generally accepted the findings of the analyses conducted by the developers, the Committee agreed that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee recognized the impact that social risk can have on cost and resource use measures and encourages measure developers to test the impact of additional social risk variables. The Committee also encouraged exploration of the impact of community-level variables. However, the Committee generally agreed that the risk-adjustment method used in these measures met the NQF criteria given the data available to the developer and the measure testing results presented.

Attribution

Public commenters raised concerns about the attribution model of NQF #2158. Commenters noted that post-acute spending drives the majority of the variation in the measures, and commenters questioned if hospitals are able to meaningfully influence their results.

The Committee had in-depth conversations on the attribution of NQF #2158. The Committee recognized that hospitals may not have complete control over the spending captured by the measure. However, the Committee believed that there are actions hospitals can take to improve their performance on the measure. Additionally, the Committee noted the need for attribution models that support care coordination and team-based care as the system aims to transition from fee-for-service to population-based payment.

Concerns About Populations Included in the Measures

Several commenters raised concerns about populations included in the measures, noting that spending can vary significantly for certain provider types and patient groups. One commenter asked for clarification on how all three measures address cancer patients. The commenter noted that there can be significant variation in treatment needs, comorbidities, and patient preferences that can influence cost and resource use.

One commenter expressed concern with the inclusion of all obstetrician-gynecologists and pharmacy resources in measures #1598 and #1604. They noted that nongeneralist obstetrician-gynecologists provide specialty care and suggested only including generalists in these two measures. The commenter also noted that providers do not control insurer formularies and that information on the cost of pharmaceuticals is not available

Support for Measures

Seven of the comments received supported the measures and agreed with the Committee's decision to recommend continued endorsement.

- Measure #1598 *Total Resource Use Population-Based PMPM Index* received two supportive comments.
- Measure #1604 *Total Cost of Care Population-Based PMPM Index* received three supportive comments.
- Measure #2158 *Medicare Spending Per Beneficiary (MSPB) – Hospital* received one supportive comment.

In addition, one general comment noted the gap in measures in this area and supported the continued endorsement of these three measures.

Measure-Specific Comments

Comments specific to particular measures, along with Committee, NQF, and developer responses, are in Appendix A.

References

[1] Martin AB, Hartman M, Washington B, et al. National health spending: faster growth in 2015 as coverage expands and utilization increases. *Health Aff (Millwood)*. 2017;36(1):166-176.

[2] Martin AB, Hartman M, Washington B, et al. National health spending: faster growth in 2015 as coverage expands and utilization increases. *Health Aff (Millwood)*. 2017;36(1):166-176.

[3] Squires D, Anderson C. *U.S. Health Care from a Global Perspective: Spending, Use of Services, Prices, and Health in 13 Countries*. *Issues in International Health Policy.* New York: The Commonwealth Fund; 2015. Available at http://www.commonwealthfund.org/~/media/files/publications/issue-brief/2015/oct/1819_squires_us_hlt_care_global_perspective_oecd_intl_brief_v3.pdf. Last accessed March 2017.

[4] Davis K, Stremikis K, Squires D, et al. *Mirror, Mirror on the Wall, How the Performance of the U.S. Health Care System Compares Internationally*. New York: The Commonwealth Fund; 2014. Available at http://www.commonwealthfund.org/~/media/files/publications/fund-report/2014/jun/1755_davis_mirror_mirror_2014.pdf. Last accessed March 2017.

[5] Martin AB, Hartman M, Washington B, et al. National health spending: faster growth in 2015 as coverage expands and utilization increases. *Health Aff (Millwood).* 2017;36(1):166-176.

[6] Martin AB, Hartman M, Washington B, et al. National health spending: faster growth in 2015 as coverage expands and utilization increases. *Health Aff (Millwood).* 2017;36(1):166-176.

[7] Squires D, Anderson C. *U.S. Health Care from a Global Perspective: Spending, Use of Services, Prices, and Health in 13 Countries*. *Issues in International Health Policy.* New York: The Commonwealth Fund; 2015. Available at http://www.commonwealthfund.org/~/media/files/publications/issue-brief/2015/oct/1819_squires_us_hlt_care_global_perspective_oecd_intl_brief_v3.pdf. Last accessed March 2017.

[8] Davis K, Stremikis K, Squires D, et al. *Mirror, Mirror on the Wall, How the Performance of the U.S. Health Care System Compares Internationally*. New York: The Commonwealth Fund; 2014. Available at http://www.commonwealthfund.org/~/media/files/publications/fund-report/2014/jun/1755_davis_mirror_mirror_2014.pdf. Last accessed March 2017.

[9] National Quality Forum (NQF). *Endorsing Cost and Resource Use Measures: Technical Report*. Washington, DC: NQF; 2014. Available at http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=74591. Last accessed March 2017.

[10] NQF. *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care*. Washington, DC: NQF; 2009. Available at http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx. Last accessed July 2017.

[11] Centers for Medicare & Medicaid Services (CMS). *National Health Expenditures 2015 Highlights*. Baltimore, MD: CMS; 2015. Available at https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/highlights.pdf. Last accessed July 2017.

[12] CMS. *National Health Expenditures 2015 Highlights*. Baltimore, MD: CMS; 2015. Available at https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/highlights.pdf. Last accessed July 2017.

Appendix A: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable; **Y**=Yes; **N**=No

Endorsed Measures

1598 Total Resource Use Population-Based PMPM Index

Submission | Specifications

Description: The Resource Use Index (RUI) is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.

A Resource Use Index when viewed together with the Total Cost of Care measure (NQF-endorsed #1604) provides a more complete picture of population based drivers of health care costs.

Numerator Statement: The numerator is calculated as the sum of (Total Medical TCRRV / Medical Member Months) + (Total Pharmacy TCRRV / Pharmacy Member Months).

Denominator Statement: The denominator is the Johns Hopkins Adjusted Clinical Grouper (ACG) risk score.

Exclusions: 1. Members over age 64, 2. Members under age 1, 3. Member enrollment less than 9 months during the one year measurement time window, 4. Members not attributed to a primary care provider, 5. Dollars per member above $125,000 are excluded (i.e. truncated)

Adjustment/Stratification: The Total Resource Use measure uses the Johns Hopkins Adjusted Clinical Grouper (ACG) which adjusts for variation in risk profile using age, gender, and diagnosis (clinical risk adjustment). The measure is also limited by insurance coverage to commercial only.

The ACG System is a statistically valid and broadly adopted risk grouper in both academic and non-academic settings with methodology derived from diagnosis information.

The ACG System assigns International Classification of Disease (ICD) diagnosis codes to 32 diagnosis groups – Aggregated Diagnosis Groups (ADGs). The assignment method is included in the ACG software for all codes. Diagnosis codes mapped to a given ADG are clinically similar and have similar expected need for healthcare resources. The assignment criteria is based on features of a condition that help predict duration and intensity of resource use. Five clinical criteria are used to determine assignment of codes: duration, severity, diagnostic certainty, type of etiology, and expected need for specialty care.

Adjusted Clinical Group actuarial cells (ACGs) build off of the ADG assignment logic described and are used to determine the morbidity profile of patient populations to more fairly assess provider performance and allow for equitable comparisons of utilization and outcomes. ACGs are defined by morbidity, age, and sex and are person-focused to categorize patients' illnesses. Based on the pattern of morbidities, the ACG approach assigns each individual to a single ACG category.

After applying measure criteria, which includes limitation to commercial only and clinical risk adjustment, socioeconomic testing was conducted that considered income and education status as potential factors beyond those already adjusted for.

Level of Analysis: Population: Community, County or City, Clinician: Group/Practice

Setting of Care: Hospital: Acute Care Facility, Ambulatory Surgery Center, Birthing Center, Clinician Office/Clinic, Hospital: Critical Care, Dialysis Facility, Emergency Department, Emergency Medical

Services/Ambulance, Home Health, Hospice, Hospital, Imaging Facility, Behavioral Health: Inpatient, Inpatient Rehabilitation Facility, Laboratory, Long Term Acute Care, Nursing Home / SNF, Other, Behavioral Health: Outpatient, Outpatient Rehabilitation, Pharmacy, Urgent Care - Ambulatory

Type of Measure: Cost/Resource Use

Data Source: Claims (Only)

Measure Steward: HealthPartners

STANDING COMMITTEE MEETING [03/15/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a.High Priority, 1b. Performance Gap, 1c. Measure Intent)

1a. High Priority: **H-16**; **M-2**; **L-1**; **I-0**; 1b. Performance Gap: **H-8**; **M-10**; **L-1**; **I-0;** Measure Intent: **H-12; M-6: L-1: I-0**

Rationale:

- To demonstrate the importance of a resource use measure, the developers cite data demonstrating healthcare spending constitutes a high proportion (17%) of the United States gross domestic product (GDP) and high healthcare costs contributes to adults forgoing healthcare. The developers suggest that this measure can support a comprehensive measurement system to identify areas of overuse.
- The developer provided performance data from 2015 dates of service from the multi-stakeholder community collaborative, Minnesota Community Measurement (MNCM) that measured the Total Resource Use of 257 provider groups, representing 1.5 million patients receiving care. MNCM found that risk-adjusted medical group resource use had variation up to 55 percent, from 22% below the state average to 33% above the state average.
- The intent of this measure is to allow measure implementers to better understand and measure overuse and underuse to drive person-centered management and accountability. A population-based measure complements condition and episode-based measures for a complete view of utilization across the measurement year.
- The Committee agreed that the measure addresses a high priority area stating that cost and affordability is a major concern in the healthcare system. It contributes to the number of uninsured, budget deficits, and medical bankruptcy. Committee members noted that understanding the total resource use is crucial to understanding how to effectively lower costs without decreasing quality.
- Committee members raised a few concerns with this measure, including whether it is possible to benchmark across multiple systems for multiple providers of the same specialty/field. HealthPartners provides a dashboard of results, which includes the measure and companion measures. HealthPartners works with providers to benchmark their performance to the plan's average performance.
- Committee members also requested information on whether mapping tools to concurrently examine their outcome measures and quality measures existed. The developer responded that they offer transparency on their website by offering both quality and experience scores for consumers to use, as well as pairing that information with overall cost information. Because the literature demonstrates that there is no direct correlation between cost and quality, the developers have not developed specifications for a joint cost and quality measure.
- A Committee member questioned how looking at medical group variability from year to year is adjusted since the measure has a relative score and groups may be improving. The developer

responded that they always index performance to the current year in order to understand where any level/unit of analysis is performing relative to the current performance of peers. In addition, there is the capability to index the previous two years to the current year, in order to show how performance trends over time.

---

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-10**; **M-7**; **L-1**; **I-1** 2b. Validity: **H-2**; **M-14**; **L-2**; **I-0**

<u>Rationale</u>:

- This per capita (population- or patient-based) measure calculates total resource use associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services and is expressed as a ratio.
- To interpret, a score greater than 1.00 indicates higher risk adjusted resource use, compared to a peer group average; a score less than 1.00 indicates less risk adjusted resource use, compared to a peer group average.
- The developer defines peer groups as a group of members, providers, geographic regions or any grouping of member data. The resource use measure will return a value that will be relative to the peer group average (e.g., 1.10 = 10% higher than the peer group average).
- The numerator is calculated as the sum of (Total Medical TCRRV / Medical Member Months) + (Total Pharmacy TCRRV / Pharmacy Member Months). The Johns Hopkins Adjusted Clinical Grouper (ACG) risk score is the measure's denominator.
- To demonstrate **measure score reliability**, the developer compared actual measure scores to scores calculated by two sampling methods:
    - Bootstrapping
    - A 90% random sample
    - The variances from Actual RUI ranged from -0.0036 to 0.0065 in the bootstrap to -0.0020 to 0.0015 in the 90% sample.
- Some Committee members expressed concerns with the reliability testing, noting concerns about the attribution approach used in the testing of the measure. The developer responded they used the attribution method used within their health plans.
- The Committee was also concerned that testing only occurred in a localized area (with data from one payer and limited geographic area), raising concerns that the results may not be generalizable and not applicable on a nationwide scale. However, this measure is currently widely used across the country and has demonstrated reliability among other users.
- Ultimately, the Committee agreed the measure met the reliability criterion.
- For this maintenance submission, the developer summarized updated validity testing conducted using provider data from 2014 and 2015. The validity and reliability testing of the measures was conducted with HealthPartners' commercial population of 470,000 members. This updated validity testing consisted of correlations of the measure components (i.e., ACG scores, unadjusted costs) and measure score with other markers of utilization.
- The developers tested the validity of the underlying data elements and performed empirical validity testing of the measure score.
- To demonstrate **data element validity**, the developer conducted a series of correlation analyses:

    - Measure components (i.e., ACG scores & Non-Risk Adjusted Total Cost Relative Resource Values (TCRRVs))
        - ACG Risk-adjusted Total Cost Index (i.e., the measure score)
        - ACG risk-adjusted Resource Use Index (RUI) (i.e., measure 1598)
        - Non-risk adjusted Total Cost Relative Resource Values (TCRRVs)
        - Price
    - Measure component - Non-Risk Adjusted TCRRVs with non-risk adjusted rates of utilization:
        - Inpatient Admits per 1,000
        - ER per 1,000
        - Outpatient surgery per 1,000
        - High Tech Radiology per 1,000
        - E&Ms per 1,000
        - Lab/Path per 1,000
        - Standard radiology per 1,000
        - Pharmacy per 1,000
        - Measure Components with Composite Utilization
- The developer noted that there is high correlation of the measure components to one another and each component's correlation with the non-risk adjusted TCRRVs as sufficient evidence for the validity of the measure components.
    - The correlation between the non-risk adjusted PMPM and the ACG risk adjusted RUI is 0.45.
- The developer attributes the low correlation between ACG and Price to fact that ACG is an estimate of expected resource use whereas price is the unit cost of services actually provided
- To demonstrate **measure score validity**, the developer conducted a series of correlation analyses:
    - ACG Risk-adjusted Risk Use Index (i.e., the measure score) with:
        - Hospital based Total Cost of Care Index
        - Professional Total Cost of Care Index
        - Pharmacy Total Cost of Care Index
        - ACG risk-adjusted Total Cost Index (i.e., measure 1604)
        - Total Price
    - Service Category RUI (i.e., Inpatient, Outpatient, Professional, Pharmacy) with risk-adjusted service category metrics:
        - Inpatient admit rate
        - ER count
        - Outpatient surgery
        - High tech Radiation
        - E&M Visits
        - Lab/Path
        - Standard Radiology
        - Prescription (Rx) Count
    - Measure Score with Composite Utilization
    - Measure Score Over Time
- The risk adjustment approach utilized in the measure is the Johns Hopkins Adjusted Clinical Grouper (ACG) method, which adjusts for age, gender, and diagnosis (i.e., clinical risk). A conceptual rationale for this risk adjustment approach is provided.
- The risk adjustment approach involves:

    - Grouping International Classification Diagnosis (ICD) diagnosis codes into 32 diagnosis groups (i.e., Aggregated Diagnosis Groups (ADGs)). These ADGs are clinically similar and expected to have similar need for healthcare resources.
    - Adjusted Clinical Groups (ACGs) are created from the ADG assignments and are defined by morbidity, age, and sex. Individual members are then assigned to a single ACG category, which quantifies their risk.
- Individual member ACG weights: Individuals are assigned to an ACG actuarial cell that has a corresponding weight reflecting relative illness burden. The ACG weight is then multiple by their number of eligible member months.
- Providers' ACG Scores are calculated as the sum of their attributed members ACG weights.
- To examine the impact of SDS on the measure scores, the developers used two measures of income – 1) tract-level income, obtained from U.S. Census Tract data, and 2) household-level, obtained from a commercially licensed consumer database purchased by HealthPartners.
- Two multiple linear regression equations were analyzed:
    - Equation 1: Tract-level income, ACG risk score, and insurance product (i.e., Commercial vs Medicaid) were regressed on total reimbursed amount per member per month; and
    - Equation 2: Household-level income, ACG risk score, and insurance product (i.e., Commercial vs Medicaid) were regressed on total reimbursed amount per member per month
- Results from both Census tract-level and household-level data sources show that income did not significantly impact the measure scores after risk adjusting for age, gender, and clinical risk, and stratifying by insurance type. The ACG score and the insurance type were determined to have a significant impact on the cost and resource use measures' variation, while income had no discernible impact. The developer hypothesized that most of the variation related to income was absorbed by variables such as medical complexity and insurance type.
- The Committee raised questions about the face validity testing. The developer responded they make their total cost of care and resource use results available to provider networks for them to review and vet. The developer has a 45-day comment period for providers to review results. There are also frequent internal meetings with medical directors and with providers to review results. In addition, a multistakeholder committee provided input on the measure.
- Ultimately, the Committee agreed the measure met the validity criterion.

---

3. Feasibility: H-14; M-4; L-0; I-0

(3a. Data is readily available; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The measure is constructed using administrative health claims, which are routinely collected and do not create undue burden for measure implementers.
- All data elements are available in defined fields within electronic sources.
- The measure uses a proprietary ACG-Johns Hopkins risk adjustment methodology. There is a cost associated with using the software required to implement the risk adjustment methodology. The developer noted that some communities have implemented the measure using different risk adjustment methodology.

- Some Committee members noted that the testing data is from Wisconsin and Minnesota, and were concerned the data is not widely generalizable. However, the Committee agreed that the measure is currently in use and that it has been feasible to implement.

---

4. Usability and Use: H-13; M-5; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The developer states that this measure is used in multiple accountability programs, including:
    - 3 Public reporting programs
    - 1 Payment program
    - 1 Public Health/Disease Surveillance program
    - 5 Quality Improvement with Benchmarking programs (external benchmarking to organizations)
    - Several Quality Improvement with Benchmarking (internal to the specific organization) programs
- The developer also cited measure page views at the National Quality Measures Clearinghouse (NQMC) from Agency for Healthcare Research and Quality (AHRQ)
    - Reported the following usage between 3/1/15 – 2/29/16
        - 5,815 page views for the Total Cost of Care Measure
        - 1,493 page views for the Total Resource Measure
- A large number of those who have adopted the measure have seen improvement due to increased transparency.
- Committee members questioned the categorization of obstetricians and gynecologists as primary care providers and noted that their patterns of resource use may differ from the other types of providers assessed by the measure. The Committee noted the need for measure users to ensure appropriate comparison groups to address this concern.
- Because the measure is disaggregated by service type provided, the Committee agreed it can be used to identify areas of improvement.

---

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-18**; **N-0**

Rationale

- The Committee agreed the measure meets the criteria, and voted to recommend the measure for continued endorsement.

---

6. Public and Member Comment: April 20, 2017 – May 19, 2017

Comments Received

- This measure received five comments. Two supported continued endorsement, and one comment was about the inclusion of non-generalist OB/GYNs and pharmaceutical resources. The remaining two comments focus on concerns around the measure's testing and usability in

states outside of those two the measure was tested in. The commenters requested more detail on the measure's current performance and implementation experiences, given that it has been in use for five years. Specific concerns raised include unintended consequences, standardized prices, a risk adjustment approach, and acceptable sample sizes. There are also additional comments focusing on concerns with SDS. One comment was received on the general draft report requesting clarification on how all three measures address cancer patients. The commenter noted that there can be significant variation in treatment needs, comorbidities, and patient preferences that can influence cost and resource use.

Developer Responses

Response 1: Testing and SDS Concerns

HealthPartners thanks the American Medical Association (AMA) for sharing its comments.

To address the AMA's first comment regarding standardized pricing, the Total Resource Use measure uses the Total Care Relative Resource Values (TCRRVs). TCRRVs are a grand linear scale of relative values designed to evaluate resource use across all types of medical services, procedures and places of service. TCRRVs are based on industry standard weighting systems (RVU, DRG, APC). The values are independent of price and can be used to evaluate providers, hospitals, physicians and health plans against their peers on their efficiency of resource use in treating like conditions.

The TCRRVs are applied at the procedure level for each component of care with the exception of inpatient, which is applied at the full admission level. There is a TCRRV lookup table for each component of care where each claim's procedure is matched with the corresponding value. The TCRRV weights that are applied to the claim is tested for accuracy and a total TCRRV is calculated.

Details regarding standardized prices can be found under section S.9.6. "Costing Method" within the measure submission form. The detailed development of the TCRRV methodology is described in a technical white paper publicly available on HealthPartners' Total Cost of Care website.

https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/cntrb_039627.pdf

The full TCRRV tables are available via our website and licensed, free of charge at: www.healthpartners.com/tcoc. Below is a sample TCRRV table:

https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/entry_188112.pdf

In regards to the comment shared about the risk adjustment approach, HealthPartners' Total Cost of Care and Resource Use measures are specified for use of the Johns Hopkins' Adjusted Clinical Groups (ACG System). However, we recognize the practicality of communities and users who have financially invested in different risk adjustment groupers. Users opting to use different risk adjuster for their analysis should test for both reliability and validity of the measures. Additionally, for comparability of results across different users, each user must use the same risk adjustment tool.

The Society of Actuaries Accuracy of Claims-Based Risk Scoring Models (2016) findings that suggest other commercially available risk groupers perform similarly.

https://www.soa.org/Files/Research/research-2016-accuracy-claims-based-risk-scoring-models.pdf

To clarify the testing was performed on provider groups with a minimum of 600 members for both Total Cost of Care and Total Resource Use measures. While HealthPartners performed the testing at the provider group level the unit of analysis can be applied to a variety of units, such as the health plan, employer group, provider group, clinic, physician or geographical area. The measures' constructs remain constant and are not dependent on level of analysis.

The HealthPartners measures of Total Cost of Care and Total Resource Use are not measures of value, they are measures that represent the affordability arm of the Triple Aim. HealthPartners is focused on the Triple Aim, quality, experience and affordability of health care to provide value for our patients. We believe it is essential when measuring affordability to complement it with quality and experience measurement, which allows members and patients to make their own value determination. The majority of the contracted provider groups in our health plan network are focusing on the same work, each having their own process improvement plan. HealthPartners' health plan implemented a Triple Aim risk sharing program over 5 years ago with Total Cost of Care representing the affordability component of the Triple Aim. To date, tens of millions of dollars have been paid out by the health plan and self-insured employer groups. Essentia Health, CentraCare Health and Fairview Health Services are provider groups outside the HealthPartners family of providers who have shared their success stories through letters of support in the usability section of the National Quality Forum measure submission process.

In addition, HealthPartners compares its own family of provider groups against the health plan network wide benchmark to identify total cost and resource use improvement opportunities. While working towards better affordability of care and reducing costs, we are still able to maintain the highest quality of care. HealthPartners' family of providers were recently reported as three of the highest performers statewide in most quality measures by Minnesota Community Measurement.

In response to the question raised about comparisons across different medical specialties, to clarify, the measure is specified as a full population measure, including all care, from all provider specialties. The unit of analysis, or attribution, is a measurement guideline for both Total Cost of Care and Total Resource Use measures. The risk adjustment accounts for variation in age, gender and the clinical risk of patients treated by various specialists (e.g. typically pediatric patients receive lower risk scores).

Minnesota Community Measurement (MNCM) is a community collaborative organization that measures Total Cost of Care and Total Resource Use annually, according to the NQF-endorsed specifications, for all provider groups in the state of Minnesota and bordering communities to drive improvement by showcasing variation through transparency. The Network for Regional Healthcare Improvement (NRHI) represents more than 30 Regional Health Improvement Collaboratives (RHICs) across the United States and three state- affiliated partners. With funding from the Robert Wood Johnson Foundation, five pioneering regional health improvement collaboratives (RHICs) are now joined by six additional regions to standardize how they report cost information. NRHI is driving a national effort to make care affordable by using the NQF-endorsed specifications to make cost and resource information consistent and transparent. Both MNCM (third year) and NRHI (first year) results were included in the usability portion of our submission documents for the committee's review.

You can't improve what you don't measure. The uptake of these measures across the country and provider engagement are the first steps to reaching our nation's goal of providing affordable care for our patients.

To address the AMA's last concern about testing sociodemographic (SDS) factors, the Total Cost of Care and Total Resource Use measures are risk adjusted for age, gender, and clinical risk profile based on diagnosis. The measures are also specified for the commercially insured population. Income and education status were explored as potential socioeconomic (SES) variables for additional adjustment due to their conceptual alignment, along with their likely data availability. Income has been viewed as a main contributor to healthcare access and affordability along with education influencing a patient's approach to the healthcare system1,2,3. Income is a continuous and granular variable. Education status is a categorical variable and difficult to create an average or median. Because income and education have been found to be correlated and because income was a more continuous and granular variable HealthPartners focused the analysis on income.

Testing was done on a data element reasonably available to HealthPartners or other users, which would not include the majority of factors listed by the AMA. HealthPartners used two separate data sources to evaluate income. The first was U.S. Census Tracts. The second was a more robust commercially licensed data source that HealthPartners has access to for other business purposes, which provided us with household level income.

To ensure the study population included lower income ranges, HealthPartners Medicaid population was included along with the full commercial book of business for testing. The Medicaid population has a different reimbursement rate (typically significantly lower) than the commercial reimbursement rate, which would result in a lower total cost of care. The Medicaid population was included to prove that the product delineation between Medicaid and Commercial sufficiently controls for the variation in cost and resource use, therefore, adding income in the model resulted in no additional explanatory power.

As stated in the SES testing analysis, after risk adjusting for age, gender, and clinical risk, and limiting by commercial product, income did not significantly impact a patient's total cost or resource use. There was less than a 1% change in performance for all provider groups when income was introduced into the model for both measures when using Census Tract data and less than a 0.5% change when using the commercially licensed data source with more granular income data.

Citations:

1. Alter D, et. al. Lesson From Canada's Universal Care: Socially Disadvantaged Patients Use More Health Services, Still Have Poorer Health.. Health Affairs doi: 10.1377/hlthaff.2009.0669 Health Aff February 2011 vol. 30 no. 2 274-283. http://content.healthaffairs.org/content/30/2/274.abstract?sid=94d288f0-331d-469e-8c11-023b272bed92

2. Lemstra, M, et.al. High health care utilization and costs associated with lower socio-economic status: results from a linked data set. Can J Public Health. 2009 May-Jun;100(3):180-3. https://www.jstor.org/stable/41995241?seq=1#page_scan_tab_contents 4.

3. United States Department of Labor, Bureau of Labor Statistics 2015. Earnings and Unemployment by Educational Attainment Status. Last Modified March 15, 2016. http://www.bls.gov/emp/ep_chart_001.htm

Response 2: Inclusion of OB/GYNs and Pharmaceutical Resources

HealthPartners thanks the American College of Obstetricians and Gynecologists (the College) for sharing its comment. The intent of the Total Cost of Care measure is to measure a provider's risk adjusted cost effectiveness at managing the population they care for. Similarly, the Total Resource Use measure is a risk adjusted measure of the frequency and intensity of services

utilized to manage a provider's patients. While all costs and resources associated with treating patients are included for evaluation, implementation of the measures and how results are used and reported are decisions that need to be considered and defined by the users.

Measure reporting guidelines (guidelines are not a part of the specifications), which include attribution methodology, have been shared to assist with implementation of the measures and appropriate comparisons across specified reporting entities. For comparability purposes, the attribution method used in Total Cost of Care and Total Resource Use measurement must be consistently applied across the population measured. In addition, a peer group or benchmark must be defined. Users must determine both the method of attribution and the peer group to be used with their own market and specific business needs in mind.

HealthPartners' attribution process, which has been vetted and accepted locally for use in our market, includes the following specialties: family medicine, internal medicine, pediatrics, geriatrics, obstetrics and gynecology (OB/GYN). We agree with the College and recognize that subspecialties and specific areas of care may not reflect primary care services. HealthPartners' measurement approach excludes specific OB/GYN specialties from attribution when measuring primary care providers (e.g. Gynecological Surgery). The measures are however versatile and could evaluate subspecialties if the peer group was limited to the subspecialty being evaluated. It is up to the user to ensure the intended use of the measure aligns with the providers being measured and is reflected in the peer group.

The measures are population-based, patient-centered and cross all categories of health care services, including pharmacy. Pharmacy contributes an estimated 20% of the total costs and resources and are driving steep trends. Therefore it is imperative to include pharmacy costs when measuring total cost of care and resource use. At HealthPartners we have and are enhancing our approaches to make the costs of drugs available to providers and consumers so the decision-making process can be fully informed. We would encourage others to do the same. While a provider may not have control over formulary drug lists, providers do have an opportunity to help educate patients on alternative drug options when there is clinical equivalence yet a large cost difference. Alternative therapies, generics or less expensive brand drugs may be options for patients that providers can help coordinate, leading to lower overall costs, lower out of pocket costs for patients, increased patient experience and most importantly, a better chance consumers can afford to fill and take their medicine as prescribed. Providers also have the ability to manage potential overuse of medications which not only reduces costs, but also improves quality of care for patients.

Response 3: Inclusion of Cancer Patients

HealthPartners' agrees that cost measures are not a marker of quality, and should not be used to draw conclusions on quality.

As a provider of cancer care, (American Society of Clinical Oncology-certified practices HealthPartners Regions Hospital Cancer Care Center and Park Nicollet Frauenshuh Cancer Center), HealthPartners' understands the complexity of cancer diagnosis, the importance of early detection and the variation in treatments as we care for cancer patients in our medical groups.

We are sensitive to the complexity of including cancer patients in our full population measure and through our testing believe the measures sufficiently adjust for cancer patients in the population through the clinical risk adjustment process and application of the measure criteria. John's Hopkins ACGs accurately and reliably adjusts for the clinical risk of a population including the risk of cancer patients and this is a primary reason why HealthPartners recommends the use

of a commercially available clinical risk adjuster rather than less effective open source adjusters. The measure criteria of a minimum of 9 month of enrollment ensures there is enough patient history to accurately assess risk and cost. In addition, the costs are truncated at $125,000 so no one patient can overly impact the performance of the measures. The peer group is also of vital importance when performing cost and resource use evaluations.

Committee Responses

The Committee recognizes the need to ensure NQF-endorsed cost and resource use measures are reliable and valid. The Committee noted that #1598 and #1604 have been widely implemented and users have supported the usefulness of the information generated by the measures.

The Committee agrees that consideration of social risk factors in risk adjustment models is a critical issue in measurement science. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer. The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and social risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between social risk factors and cost and resource use.

While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee recognized the impact that social risk can have on cost and resource use measures and encourages measure developers to test the impact of additional social risk variables. The Committee also encouraged exploration of the impact of community-level variables. However, the Committee generally agreed that the risk adjustment method used in these measures met the NQF criteria given the data available to the developer, and the measure testing results presented.

After reviewing the comments received, and the developer's response, the Committee does not wish to reconsider its recommendations on any of the three measures. While the specific attribution model is currently outside of the endorsement criteria, the Committee recognizes the need for further testing of attribution models.

---

7. Consensus Standards Approval Committee (CSAC) Vote: Y-13; N-0

- **Decision**: Approved for continued endorsement

---

8. Appeals

No appeals received.

1604 Total Cost of Care Population-Based PMPM Index

Submission | Specifications

Description: Total Cost of Care reflects a mix of complicated factors such as patient illness burden, service utilization and negotiated prices. Total Cost Index (TCI) is a measure of a primary care provider's risk adjusted cost effectiveness at managing the population they care for. TCI includes all costs associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.

A Total Cost Index when viewed together with the Total Resource Use measure (NQF-endorsed #1598) provides a more complete picture of population based drivers of health care costs.

Numerator Statement: The numerator is calculated as the sum of (Total Medical Cost/Medical Member Months) + (Total Pharmacy Cost/Pharmacy Member Months)

Denominator Statement: The denominator is the Johns Hopkins Adjusted Clinical Grouper (ACG) risk score

Exclusions: 1. Members over age 64, 2. Members under age 1, 3. Member enrollment less than 9 months during the one year measurement time window, 4. Members not attributed to a primary care provider, 5. Dollars per member above $125,000 are excluded (i.e. truncated)

Adjustment/Stratification:

The Total Cost of Care measure uses the Johns Hopkins Adjusted Clinical Grouper (ACG) which adjusts for variation in risk profile using age, gender, and diagnosis (clinical risk adjustment). The measure is also limited by insurance coverage to commercial only.

The ACG System is a statistically valid and broadly adopted risk grouper in both academic and non-academic settings with methodology derived from diagnosis information.

The ACG System assigns International Classification of Disease (ICD) diagnosis codes to 32 diagnosis groups – Aggregated Diagnosis Groups (ADGs). The assignment method is included in the ACG software for all codes. Diagnosis codes mapped to a given ADG are clinically similar and have similar expected need for healthcare resources. The assignment criteria is based on features of a condition that help predict duration and intensity of resource use. Five clinical criteria are used to determine assignment of codes: duration, severity, diagnostic certainty, type of etiology, and expected need for specialty care.

Adjusted Clinical Group actuarial cells (ACGs) build off of the ADG assignment logic described and are used to determine the morbidity profile of patient populations to more fairly assess provider performance and allow for equitable comparisons of utilization and outcomes. ACGs are defined by morbidity, age, and sex and are person-focused to categorize patients' illnesses. Based on the pattern of morbidities, the ACG approach assigns each individual to a single ACG category.

After applying measure criteria, which includes limitation to commercial only and clinical risk adjustment, socioeconomic testing was conducted that considered income and education status as potential factors beyond those already adjusted for.

Level of Analysis: Population : Community, County or City, Clinician : Group/Practice

Setting of Care: Hospital : Acute Care Facility, Ambulatory Surgery Center, Birthing Center, Clinician Office/Clinic, Hospital : Critical Care, Dialysis Facility, Emergency Department, Emergency Medical Services/Ambulance, Home Health, Hospice, Hospital, Imaging Facility, Behavioral Health : Inpatient, Inpatient Rehabilitation Facility, Laboratory, Long Term Acute Care, Nursing Home / SNF, Other, Behavioral Health : Outpatient, Outpatient Rehabilitation, Pharmacy, Urgent Care - Ambulatory

Type of Measure: Cost/Resource Use

Data Source: Claims (Only)

Measure Steward: HealthPartners

STANDING COMMITTEE MEETING [03/15/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. High Priority, 1b. Performance Gap, 1c. Measure Intent)

1a. High Priority: **H-16**; **M-2**; **L-1**; **I-0**; 1b. Performance Gap: **H-8**; **M-10**; **L-1**; **I-0**; 1c. Measure Intent: **H-12**; **M-6**; **L-1**; **I-0**

Rationale:

- To demonstrate the importance of measuring cost, the developers cite data demonstrating healthcare spending constitutes a high proportion (17%) of the United States gross domestic product (GDP) and high healthcare costs contributes to adults forgoing healthcare.
- The developers suggest that this measure can support a comprehensive measurement system to identify areas of overuse.
- The developer presents performance data from 2015 dates of service from the multi-stakeholder community collaborative, Minnesota Community Measurement (MNCM) measured the Total Resource Use of 257 provider groups, representing 1.5 million patients receiving care. The 2015 risk-adjusted total cost of care per member per month on average was $474, with a range of $365 to $916. Eighty percent of provider groups were between $394 and $555 per member per month. The developer did not provide data on changes in performance over time.
- To examine differences in measure scores by age and gender, the developer examined the distribution of scores in single specialty obstetric and pediatric groups. Data from these analyses were not provided, but the developer states scores were uniformly distributed and not clustered.
- The intent of this measure is to allow measure implementers to better understand and measure overuse and underuse to drive person-centered management and accountability. A population-based measure complements conditions and episode-based measure for a complete view of utilization across the measurement year.
- Due to similarity between #1598 and #1604 in the measure structure and logic, the Standing Committee agreed to apply the votes on the #1598 Importance criteria to #1604.

---

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-10**; **M-8**; **L-0**; **I-0** 2b. Validity: **H-3**; **M-14**; **L-1**; **I-0**

Rationale:

- This per capita (population- or patient-based) measure calculates the total cost of care of a commercial population and is expressed as a ratio.
- To interpret, a score greater than 1.00 indicates a higher paid risk adjusted PMPM value, compared to a peer group average; a score less than 1.00 indicates less paid risk adjusted PMPM value, compared to a peer group average.
- The choice of a peer group is at the discretion of the measure user and can include the internal medicine, family medicine, pediatrics, geriatrics, and OB/GYN specialties and physician,

physician assistant, and nurse practitioner provider types. The peer group's average is set at the benchmark.

- The numerator is calculated as the sum of (Total Medical Cost / Medical Member Months) + (Total Pharmacy Cost / Pharmacy Member Months).
- The Johns Hopkins Adjusted Clinical Grouper (ACG) risk scores constitutes the measure's denominator.
- The developer reported one change to the measure specifications. Previously, values were winsorized if their total medical and pharmacy costs exceeded $100,000. The developers increased this amount to $125,000 to account for the natural rise in healthcare costs over the past several years.
- For this maintenance submission, validity and reliability testing of the measures was conducted with HealthPartners' commercial population which is 470,000 members.
- Reliability testing was performed at the measure score level.
- To demonstrate measure score reliability, the developer compared actual measure scores to scores calculated by two sampling methods:
    - Bootstrapping:
        - Difference between actual score and sampling score showed a range of -0.0059 to 0.0075.
        - Results variation for within groups was <1%; Between groups >110%.
    - A 90% random sample:
        - Difference between actual score and sampling score showed a range of -0.0022 to 0.0012.
- The Committee agreed with the previous endorsement's assessment, noting that the measure continues to have a high degree of reliability. They noted that the measure's construction and calculation logic and testing results remain strong, and that the measure is in widespread use.
- The Committee highlighted the smaller variances in the reliability testing compared to the figures presented in the previous endorsement, and also noted the provider performance data was consistent between endorsement periods.
- Ultimately, the Committee agreed the measure met the reliability criterion.
- For this maintenance submission, the developer summarized updated validity testing conducted using provider data from 2014 and 2015. This updated validity testing consisted of correlations the measure components (i.e., ACG scores, unadjusted costs) and measure score with other markers of utilization.
- To demonstrate **data element validity**, the developer conducted a series of correlation analyses:
    - Measure components (i.e., ACG scores & Non-risk adjusted per member per month value (Non-Risk Adjusted PMPMs))
        - ACG Risk-adjusted Total Cost Index (i.e., the measure score)
        - ACG risk-adjusted Resource Use Index (RUI) (i.e., measure 1598)
        - Non-risk adjusted Total Cost Relative Resource Values (TCRRVs)
        - Price
    - Measure component - Non-Risk Adjusted PMPMs with non-risk adjusted rates of utilization:
        - Inpatient Admits per 1,000
        - ER per 1,000
        - Outpatient surgery per 1,000
        - High Tech Radiology per 1,000
        - E&Ms per 1,000

        - Lab/Path per 1,000
        - Standard radiology per 1,000
        - Pharmacy per 1,000
    - Measure Components with Composite Utilization
- The developer notes there is a high correlation of the measure components to one another and each component's correlation with the non-risk adjusted TCRRVs as sufficient evidence for the validity of the measure components.
- The correlation between the non-risk adjusted PMPM and the ACG Risk Adjusted TCI is 0.79.
- The developer attributes the low correlated between ACG and Price to fact that ACG is an estimate of expected resource use whereas price is the unit cost of services actually provided.
- To demonstrate **measure score validity**, the developer conducted a series of correlation analyses:
    - ACG Risk-adjusted Total Cost Index (i.e., the measure score) with:
        - Hospital based Total Cost of Care Index
        - Professional Total Cost of Care Index
        - Pharmacy Total Cost of Care Index
        - ACG risk-adjusted Resource Use Index (RUI) (i.e., measure 1598)
        - Total Price
    - Service Category TCI (i.e., Inpatient, Outpatient, Professional, Pharmacy) with risk-adjusted service category metrics:
        - Inpatient admit rate
        - ER count
        - Outpatient surgery
        - High tech Radiation
        - E&M Visits
        - Lab/Path
        - Standard Radiology
        - Prescription (Rx) Count
    - Measure Score with Composite Utilization
    - Measure Score Over time
- The risk adjustment approach utilized in the measure is the Johns Hopkins Adjusted Clinical Grouper (ACG) method, which adjusts for age, gender, and diagnosis (i.e., clinical risk). A conceptual rationale for this risk adjustment approach is provided.
- The risk adjustment approach involves:
    - Grouping International Classification Diagnosis (ICD) diagnosis codes into 32 diagnosis groups (i.e., Aggregated Diagnosis Groups (ADGs)). These ADGs are clinically similar and expected to have similar need for healthcare resources.
    - Adjusted Clinical Groups (ACGs) are created from the ADG assignments and are defined by morbidity, age, and sex. Individual members are then assigned to a single ACG category, which quantifies their risk.
- Individual member ACG weights: Individuals are assigned to an ACG actuarial cell that has a corresponding weight reflecting relative illness burden. The ACG weight is then multiple by their number of eligible member months.
- Providers' ACG Scores are calculated as the sum of their attributed members ACG weights.
- Given the ACG risk adjustment approach is owned by Johns Hopkins, the developer does not provide a summary of statistical results of the analyses conducted on ACG risk model as that information is proprietary

- To examine the impact of SDS on the measure scores, the developers used two measures of income – 1) tract-level income, obtained from U.S. Census Tract data, and 2) household-level, obtained from a commercially licensed consumer database purchased by HealthPartners.
- Two multiple linear regression equations were analyzed:
    - Equation 1: Tract-level income, ACG risk score, and insurance product (i.e., Commercial vs Medicaid) were regressed on total reimbursed amount per member per month; and
    - Equation 2: Household-level income, ACG risk score, and insurance product (i.e., Commercial vs Medicaid) were regressed on total reimbursed amount per member per month
    - Results from both Census tract-level and household-level data sources show that income did not significantly impact the measure scores after risk adjusting for age, gender, and clinical risk, and stratifying by insurance type. The ACG score and the insurance type were determined to have a significant impact on the cost and resource use measures' variation, while income had no discernible impact. The developer hypothesized that most of the variation related to income was absorbed by variables such as medical complexity and insurance type.
- The Committee asked for clarification of how price is included and how different payment models are handled in the measure. The developer clarified that #1604 is a total cost measure, which includes the plan liability plus the member liability. The measure user can select the payment system (e.g., fee-for-service or DRG-based payment).

---

3. Feasibility: H-12; M-6; L-0; I-0

(3a. Data is readily available; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The measure is constructed using administrative health claims, which are routinely collected and do not create undue burden for measure implementers.
- All data elements are available in defined fields within electronic sources.
- The measure uses a proprietary ACG-Johns Hopkins risk adjustment methodology. There is a cost associated with using the software required to implement the risk adjustment methodology. The developer noted that some communities have implemented the measure with alternate risk adjustment methodologies.
- Some Committee members noted that the testing data is from Wisconsin and Minnesota, and were concerned the data is not widely generalizable. However, the Committee agreed that the measure is currently in use and that it has been feasible to implement.

---

4. Usability and Use: H-12; M-6; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The developer states that multiple accountability programs are using this measure, including:
    - 3 Public reporting programs
    - 1 Payment program
    - 1 Public Health/Disease Surveillance program

    - 5 Quality Improvement with Benchmarking programs (external benchmarking to organizations)
    - Several Quality Improvement with Benchmarking (internal to the specific organization) programs
- The developer also cited measure page views at the National Quality Measures Clearinghouse (NQMC) from Agency for Healthcare Research and Quality (AHRQ)
    - Reported the following usage between 3/1/15 – 2/29/16
        - 5,815 page views for the Total Cost of Care Measure
        - 1,493 page views for the Total Resource Measure
- A large number of those who have adopted the measure have seen improvement due to increased transparency.
- Committee members questioned the categorization of obstetricians and gynecologists as primary care providers and noted that their patterns of resource use may differ from the other types of providers assessed by the measure. The Committee noted the need for measure users to ensure appropriate comparison groups to address this concern.
- The Committee inquired about the implementation cost when implementing measures #1598 and #1604, and how it would affect small health systems and physician groups. The developer noted that the only fee for use of the measure is the cost of the commercial risk adjuster.

---

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-18**; **N-0**

- The Committee agreed the measure met the NQF criteria and recommended it for continued endorsement.

---

6. Public and Member Comment: April 20, 2017 – May 19, 2017

Comments Received

- This measure received five comments. Three supported continued endorsement, and one requested more information regarding the inclusion of OB/GYNs. The remaining comment focuses on concerns around the measure's testing and whether the measure's specifications are precise enough to ensure consistent implementation. Specific concerns raised include unintended consequences, standardized prices, a risk adjustment approach, and acceptable sample sizes. One comment was received on the general draft report requesting clarification on how all three measures address cancer patients. The commenter noted that there can be significant variation in treatment needs, comorbidities, and patient preferences that can influence cost and resource use.

Developer Responses

Response 1: Testing and SDS concerns

HealthPartners thanks the American Medical Association (AMA) for sharing its comments.

To address the AMA's first comment regarding standardized pricing, the Total Resource Use measure uses the Total Care Relative Resource Values (TCRRVs). TCRRVs are a grand linear scale of relative values designed to evaluate resource use across all types of medical services, procedures and places of service. TCRRVs are based on industry standard weighting systems

(RVU, DRG, APC). The values are independent of price and can be used to evaluate providers, hospitals, physicians and health plans against their peers on their efficiency of resource use in treating like conditions.

The TCRRVs are applied at the procedure level for each component of care with the exception of inpatient, which is applied at the full admission level. There is a TCRRV lookup table for each component of care where each claim's procedure is matched with the corresponding value. The TCRRV weights that are applied to the claim is tested for accuracy and a total TCRRV is calculated.

Details regarding standardized prices can be found under section S.9.6. "Costing Method" within the measure submission form. The detailed development of the TCRRV methodology is described in a technical white paper publicly available on HealthPartners' Total Cost of Care website.

https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/cntrb_039627.pdf

The full TCRRV tables are available via our website and licensed, free of charge at: www.healthpartners.com/tcoc. Below is a sample TCRRV table:

https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/entry_188112.pdf

In regards to the comment shared about the risk adjustment approach, HealthPartners' Total Cost of Care and Resource Use measures are specified for use of the Johns Hopkins' Adjusted Clinical Groups (ACG System). However, we recognize the practicality of communities and users who have financially invested in different risk adjustment groupers. Users opting to use different risk adjuster for their analysis should test for both reliability and validity of the measures. Additionally, for comparability of results across different users, each user must use the same risk adjustment tool.

The Society of Actuaries Accuracy of Claims-Based Risk Scoring Models (2016) findings that suggest other commercially available risk groupers perform similarly. https://www.soa.org/Files/Research/research-2016-accuracy-claims-based-risk-scoring-models.pdf

To clarify the testing was performed on provider groups with a minimum of 600 members for both Total Cost of Care and Total Resource Use measures. While HealthPartners performed the testing at the provider group level the unit of analysis can be applied to a variety of units, such as the health plan, employer group, provider group, clinic, physician or geographical area. The measures' constructs remain constant and are not dependent on level of analysis.

The HealthPartners measures of Total Cost of Care and Total Resource Use are not measures of value, they are measures that represent the affordability arm of the Triple Aim. HealthPartners is focused on the Triple Aim, quality, experience and affordability of health care to provide value for our patients. We believe it is essential when measuring affordability to complement it with quality and experience measurement, which allows members and patients to make their own value determination. The majority of the contracted provider groups in our health plan network are focusing on the same work, each having their own process improvement plan. HealthPartners' health plan implemented a Triple Aim risk sharing program over 5 years ago with Total Cost of Care representing the affordability component of the Triple Aim. To date, tens of millions of dollars have been paid out by the health plan and self-insured employer groups. Essentia Health, CentraCare Health and Fairview Health Services are provider groups outside the

HealthPartners family of providers who have shared their success stories through letters of support in the usability section of the National Quality Forum measure submission process.

In addition, HealthPartners compares its own family of provider groups against the health plan network wide benchmark to identify total cost and resource use improvement opportunities. While working towards better affordability of care and reducing costs, we are still able to maintain the highest quality of care. HealthPartners' family of providers were recently reported as three of the highest performers statewide in most quality measures by Minnesota Community Measurement.

In response to the question raised about comparisons across different medical specialties, to clarify, the measure is specified as a full population measure, including all care, from all provider specialties. The unit of analysis, or attribution, is a measurement guideline for both Total Cost of Care and Total Resource Use measures. The risk adjustment accounts for variation in age, gender and the clinical risk of patients treated by various specialists (e.g. typically pediatric patients receive lower risk scores).

Minnesota Community Measurement (MNCM) is a community collaborative organization that measures Total Cost of Care and Total Resource Use annually, according to the NQF-endorsed specifications, for all provider groups in the state of Minnesota and bordering communities to drive improvement by showcasing variation through transparency. The Network for Regional Healthcare Improvement (NRHI) represents more than 30 Regional Health Improvement Collaboratives (RHICs) across the United States and three state- affiliated partners. With funding from the Robert Wood Johnson Foundation, five pioneering regional health improvement collaboratives (RHICs) are now joined by six additional regions to standardize how they report cost information. NRHI is driving a national effort to make care affordable by using the NQF-endorsed specifications to make cost and resource information consistent and transparent. Both MNCM (third year) and NRHI (first year) results were included in the usability portion of our submission documents for the committee's review.

You can't improve what you don't measure. The uptake of these measures across the country and provider engagement are the first steps to reaching our nation's goal of providing affordable care for our patients.

To address the AMA's last concern about testing sociodemographic (SDS) factors, the Total Cost of Care and Total Resource Use measures are risk adjusted for age, gender, and clinical risk profile based on diagnosis. The measures are also specified for the commercially insured population. Income and education status were explored as potential socioeconomic (SES) variables for additional adjustment due to their conceptual alignment, along with their likely data availability. Income has been viewed as a main contributor to healthcare access and affordability along with education influencing a patient's approach to the healthcare system1,2,3. Income is a continuous and granular variable. Education status is a categorical variable and difficult to create an average or median. Because income and education have been found to be correlated and because income was a more continuous and granular variable HealthPartners focused the analysis on income.

Testing was done on a data element reasonably available to HealthPartners or other users, which would not include the majority of factors listed by the AMA. HealthPartners used two separate data sources to evaluate income. The first was U.S. Census Tracts. The second was a more robust commercially licensed data source that HealthPartners has access to for other business purposes, which provided us with household level income.

To ensure the study population included lower income ranges, HealthPartners Medicaid population was included along with the full commercial book of business for testing. The Medicaid population has a different reimbursement rate (typically significantly lower) than the commercial reimbursement rate, which would result in a lower total cost of care. The Medicaid population was included to prove that the product delineation between Medicaid and Commercial sufficiently controls for the variation in cost and resource use, therefore, adding income in the model resulted in no additional explanatory power.

As stated in the SES testing analysis, after risk adjusting for age, gender, and clinical risk, and limiting by commercial product, income did not significantly impact a patient's total cost or resource use. There was less than a 1% change in performance for all provider groups when income was introduced into the model for both measures when using Census Tract data and less than a 0.5% change when using the commercially licensed data source with more granular income data.

Citations:

1. Alter D, et. al. Lesson From Canada's Universal Care: Socially Disadvantaged Patients Use More Health Services, Still Have Poorer Health.. Health Affairs doi: 10.1377/hlthaff.2009.0669Health Aff February 2011 vol. 30 no. 2 274-283 http://content.healthaffairs.org/content/30/2/274.abstract?sid=94d288f0-331d-469e-8c11-023b272bed92

2. Lemstra, M, et.al. High health care utilization and costs associated with lower socio-economic status: results from a linked data set. Can J Public Health. 2009 May-Jun;100(3):180-3. https://www.jstor.org/stable/41995241?seq=1#page_scan_tab_contents 4.

3. United States Department of Labor, Bureau of Labor Statistics 2015. Earnings and Unemployment by Educational Attainment Status. Last Modified March 15, 2016. http://www.bls.gov/emp/ep_chart_001.htm

Response 2: Inclusion of OB/GYNs and Pharmaceutical Resources

HealthPartners thanks the American College of Obstetricians and Gynecologists (the College) for sharing its comment. The intent of the Total Cost of Care measure is to measure a provider's risk adjusted cost effectiveness at managing the population they care for. Similarly, the Total Resource Use measure is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider's patients. While all costs and resources associated with treating patients are included for evaluation, implementation of the measures and how results are used and reported are decisions that need to be considered and defined by the users.

Measure reporting guidelines (guidelines are not a part of the specifications), which include attribution methodology, have been shared to assist with implementation of the measures and appropriate comparisons across specified reporting entities. For comparability purposes, the attribution method used in Total Cost of Care and Total Resource Use measurement must be consistently applied across the population measured. In addition, a peer group or benchmark must be defined. Users must determine both the method of attribution and the peer group to be used with their own market and specific business needs in mind.

HealthPartners' attribution process, which has been vetted and accepted locally for use in our market, includes the following specialties: family medicine, internal medicine, pediatrics, geriatrics, obstetrics and gynecology (OB/GYN). We agree with the College and recognize that subspecialties and specific areas of care may not reflect primary care services. HealthPartners' measurement approach excludes specific OB/GYN specialties from attribution when measuring primary care providers (e.g. Gynecological Surgery). The measures are however versatile and

could evaluate subspecialties if the peer group was limited to the subspecialty being evaluated. It is up to the user to ensure the intended use of the measure aligns with the providers being measured and is reflected in the peer group.

The measures are population-based, patient-centered and cross all categories of health care services, including pharmacy. Pharmacy contributes an estimated 20% of the total costs and resources and are driving steep trends. Therefore it is imperative to include pharmacy costs when measuring total cost of care and resource use. At HealthPartners we have and are enhancing our approaches to make the costs of drugs available to providers and consumers so the decision-making process can be fully informed. We would encourage others to do the same. While a provider may not have control over formulary drug lists, providers do have an opportunity to help educate patients on alternative drug options when there is clinical equivalence yet a large cost difference. Alternative therapies, generics or less expensive brand drugs may be options for patients that providers can help coordinate, leading to lower overall costs, lower out of pocket costs for patients, increased patient experience and most importantly, a better chance consumers can afford to fill and take their medicine as prescribed. Providers also have the ability to manage potential overuse of medications which not only reduces costs, but also improves quality of care for patients.

Response 3: Inclusion of Cancer Patients

HealthPartners' agrees that cost measures are not a marker of quality, and should not be used to draw conclusions on quality.

As a provider of cancer care, (American Society of Clinical Oncology-certified practices HealthPartners Regions Hospital Cancer Care Center and Park Nicollet Frauenshuh Cancer Center), HealthPartners' understands the complexity of cancer diagnosis, the importance of early detection and the variation in treatments as we care for cancer patients in our medical groups.

We are sensitive to the complexity of including cancer patients in our full population measure and through our testing believe the measures sufficiently adjust for cancer patients in the population through the clinical risk adjustment process and application of the measure criteria. John's Hopkins ACGs accurately and reliably adjusts for the clinical risk of a population including the risk of cancer patients and this is a primary reason why HealthPartners recommends the use of a commercially available clinical risk adjuster rather than less effective open source adjusters. The measure criteria of a minimum of 9 month of enrollment ensures there is enough patient history to accurately assess risk and cost. In addition, the costs are truncated at $125,000 so no one patient can overly impact the performance of the measures. The peer group is also of vital importance when performing cost and resource use evaluations.

Committee Response

The Committee recognizes the need to ensure NQF-endorsed cost and resource use measures are reliable and valid. The Committee noted that #1598 and #1604 have been widely implemented and users have supported the usefulness of the information generated by the measures.

The Committee agrees that consideration of social risk factors in risk adjustment models is a critical issue in measurement science. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer. The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and social risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to

test the conceptual and empirical relationship between social risk factors and cost and resource use.

While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee recognized the impact that social risk can have on cost and resource use measures and encourages measure developers to test the impact of additional social risk variables. The Committee also encouraged exploration of the impact of community-level variables. However, the Committee generally agreed that the risk adjustment method used in these measures met the NQF criteria given the data available to the developer, and the measure testing results presented.

After reviewing the comments received, and the developer's response, the Committee does not wish to reconsider its recommendations on any of the three measures. While the specific attribution model is currently outside of the endorsement criteria, the Committee recognizes the need for further testing of attribution models).

---

7. Consensus Standards Approval Committee (CSAC) Vote: Y-13; N-0

- **Decision**: Approved for continued endorsement

---

8. Appeals

No appeals received.

2158 Medicare Spending Per Beneficiary (MSPB) - Hospital

Submission | Specifications

Description: The Medicare Spending Per Beneficiary (MSPB) - Hospital measure evaluates hospitals' risk-adjusted episode costs relative to the risk-adjusted episode costs of the national median hospital. Specifically, the MSPB-Hospital measure assesses the cost to Medicare for services performed by hospitals and other healthcare providers during an MSPB-Hospital episode, which is comprised of the periods immediately prior to, during, and following a patient's hospital stay. The MSPB-Hospital measure is not condition specific and uses standardized prices when measuring costs. Beneficiary populations eligible for the MSPB-Hospital calculation include Medicare beneficiaries enrolled in Medicare Parts A and B who were discharged from short-term acute Inpatient Prospective Payment System (IPPS) hospitals during the period of performance.

Numerator Statement: Average spending level for the hospital's MSPB-hospital episodes divided by the average expected episode spending level for the hospital's episodes, multiplied by the average spending over all episodes across all hospitals nationally.

Denominator Statement: The episode-weighted median MSPB-Hospital amount across all episodes nationally

Exclusions: 1. Acute-to-acute transfer episodes: based on claim discharge code, 2. Death episodes: beneficiary dies during the measurement episode, 3. Overlapping episodes: occurrence of an inpatient admission during the 30 days post-discharge of an index admission is not considered a new index admission, 4. Outlier episodes: episode whose relative scores fall above the 99th percentile or below the 1st percentile of the distribution of residuals

Adjustment/Stratification: The MSPB-Hospital risk adjustment model is based on the CMS-HCC risk adjustment methodology, but unlike the CMS-HCC methodology, the MSPB-Hospital model does NOT adjust for sex. The measure employs an ordinary least squares (OLS) regression model and a separate OLS regression model to obtain the predicted episode cost for each Major Diagnostic Category that is determined by the MS-DRG of the index hospital stay. The MSPB-risk adjustment model includes indicators of age, disability status, end-stage renal disease status, long-term care, severity of illness (measured via hierarchical conditions categories (HCC)), and the MS-DRG of the index admission.

Level of Analysis: Facility

Setting of Care: Hospital: Acute Care Facility

Type of Measure: Cost/Resource Use

Data Source: Claims (Only), Other

Measure Steward: Centers for Medicare & Medicaid Services

STANDING COMMITTEE MEETING [03/15/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. High Priority, 1b. Performance Gap. 1c. Measure Intent)

1a. High Priority: **H-18**; **M-0**; **L-0**; **I-0**; 1b. Performance Gap: **H-12**; **M-6**; **L-0**; **I-0;** 1c. Measure Intent: **H-13**; **M-5**; **L-0**; **I-0;**

Rationale:

- To demonstrate this measure focuses on a high-priority area, the developers cite data indicating Medicare expenditures accounted for 3.6% ($647.6 billion) of the Gross Domestic

Product (GDP) in 2015 and hospital benefits accounted for 30% ($188.3 billion) of those Medicare expenditures. The developer also cites data indicating Medicare expenditures will account for 6.0 to 9.1% of the GDP by 2090, if current trends continue.

- The developer provided data from 2015 on performance trends for 3,298 inpatient prospective payment system hospitals. Measure scores ranged from 0.59 to 2.25 with an interquartile range of 0.09. These values indicate performance variation among providers.
- The developer states the measure's intent is to, "…incentivize hospitals to coordinate care and reduce unnecessary utilization during the period immediately prior to, during, and in the 30 days after a hospital discharge."
- The developer describes the measure construct as encompassing all types of services received (i.e., Part A and Part B claims) during the episode and states that the all-cause nature of the measure maximizes its ability to promote hospital efficiency by promoting coordination across settings and providers.
- Ultimately, the Committee agreed the measure met the Importance to Measure and Report criterion.

---

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-8**; **M-10**; **L-0**; **I-0;** 2b. Validity: **H-4**; **M-9**; **L-5**; **I-0**

<u>Rationale</u>:

- This hospital level measure calculates the ratio of payment standardized, risk-adjusted Medicare Spending Per Beneficiary (MSPB) amount for each hospital divided by the episode-weighted median MSPB-Hospital amount across all hospitals. Lower scores are better.
- The numerator includes the average actual spending level for the hospital's MSPB-hospital episodes divided by the average expected episode spending level for the hospital's episodes, multiplied by the average spending over all episodes across all hospitals nationally.
- The denominator includes the episode-weighted median MSPB-Hospital amount across all episodes nationally.
- For this maintenance submission, the developer tested data element and measure score reliability using data from approximately 5.5 million episodes that occurred between 1/1/2015 and 12/1/2015.
- The developer assessed reliability at both the data element and measure score levels.
- To demonstrate **data element reliability**, the developer cited CMS auditing and data analysis programs that regularly assess the accuracy of the claims submitted to CMS. To enhance the reliability of the data elements, the measure is calculated using data with a 3 month claims run-out from the end of the performance period.
- To demonstrate **measure score reliability**, the developer conducted two analyses:
    - Test/Retest analysis: a similar approach was used as in the initial testing, but the developer compared two random subsets from 2015, and compared the set of 2015 episodes to the set of 2014 episodes.
    - Reliability score: the developer used a similar approach to calculate reliability scores.
- Test/Retest analysis:
    - 2015 vs. 2014 measure scores: over 75% of hospitals in the lowest-spending quintile in one year were in the same quintile in the other year; over 74% of hospitals in the high-

spending quintile in one year were in the same quintile in the other year. Spearman rank correlation coefficient for a hospital across the two years was 0.85 and the Pearson correlation coefficient was 0.81, both indicating a high degree of agreement between the two years.
    - 2015 random subset$_1$ vs. 2015 random subset$_2$: over 72% of hospitals in the lowest-spending quintile in one subset were in the same quintile in the other subset; over 71% of hospitals in the highest-spending quintile in one subset were in the same quintile in the other subset. Spearman rank correlations for a hospital across samples was 0.82, and the Pearson correlation coefficient was 0.70. The developer states this lower value for the Pearson correlation coefficient is acceptable given the outcome of interest (i.e., measure scores) is identical in the two subsets and this negatively affects the calculation of the correlation coefficient.
- Reliability score calculations:
    - For hospitals with at least 25 MSPB-Hospital episodes, over 99% had a reliability score greater than 0.4 and 67.9% had a reliability score greater than 0.9. The developer cites previous work supporting 0.4 as the lower limit of moderate reliability
- The Committee agreed that overall the measure is clearly specified and can be reliably implemented.
- One Committee member questioned reliability of the disability variable because the original reason for the enrollment code from the Enrollment Database (ED) can reset when a patient reached 65 years of age, wiping out the disability code. The Committee member suggested a better disability indicator could be found in the CMS Integrated Data Repository. The developer stated they would review the recommendation.
- The Committee expressed concern over the developer's use of a 0.4 reliability threshold, stating that such a level is below commonly accepted standards. The developer responded that they had examined a higher reliability threshold of 0.7 and found that 93% of providers meet or exceeded that threshold. The developer also stated that aiming for a reliability threshold higher than 0.7 may be unrealistic as there are natural variations in spending across patients, and such variations affect the measure's reliability. The Committee found this additional explanation acceptable.
- During the reliability discussion, the Committee questioned the implications of the measure's 25 episodes minimum inclusion criterion. The developer shared additional analyses examining the effect of different number of minimum episodes on the measure's reliability, which found that significant increases in reliability were not achieved until the minimum was set near 110 episodes. The developer chose not to utilize this minimum value because it increased the number of excluded providers, and cited the need to balance increased reliability with provider inclusivity.
- The developer conducted validity testing at the measure score level. Testing results indicated the measure score was moderately correlated with Hospital Referral Regions (HRR) levels from 2007 to 2014 (range of Spearman rank correlation coefficients: 0.53-0.63; range of Pearson correlation coefficients: 0.51-0.61). The measure score was also moderately correlated with other measures of service utilization, specifically professional E&M services (Pearson correlation coefficient: 0.42) and post-acute skilled nursing services Pearson correlation coefficient: 0.52). The developer also examined cost variations by time period and found the post-discharge period accounted for 84% of the total variance in the measure score.

- The MSPB-Hospital risk adjustment model is based on the CMS-HCC risk adjustment methodology, but, unlike the CMS-HCC methodology, the MSPB-Hospital model does NOT adjust for sex.
- The measure employs an ordinary least squares (OLS) regression model and a separate OLS regression model to obtain the predicted episode cost for each Major Diagnostic Category that is determined by the MS-DRG of the index hospital stay.
- The MSPB-risk adjustment model includes indicators of age, disability status, end-stage renal disease status, long-term care, severity of illness (measured via hierarchical conditions categories (HCC)), and the MS-DRG of the index admission.
- Race (i.e., Non-Black and Black) and income-to-poverty ratio were used to examine the impact of SDS on the risk adjustment model. F-test of significance was conducted to assess the impact An F-test of significance allows one to see whether the addition of a variable to a regression model has a significant effect on the outcome variable. Both race and income-to-poverty ratio were significant predictors of the measure score, but when included in the risk adjustment regression with other variables, minor change occurred in the measure score.
- The developers stated that the minimal effect of these two variables likely indicates SDS effects on measure scores are largely captured through existing risk adjustment variables and their inclusion in the risk adjustment model is not necessary.
- The Committee raised concerns about the use of zip-code level income in the risk adjustment testing, stating that this level of income is not sensitive enough to capture individual-level attributes that might affect the measure score. The Committee was concerned that dual eligibility status was not tested, citing results of a recent report by ASPE. The developer responded by sharing the results of additional analyses examining the effect of dual eligibility status on the measure score. These results indicated inclusion of dual eligibility status was not significant for the majority of providers. The developer noted that when dual eligibility status was used in the model, more than 98% of hospitals had a change in measure score of less than 0.01 in magnitude. The Committee discussed these additional results and agreed the results were helpful, but urged to the developer to provide more information about the minority of hospitals that had a significant change in measure scores so as to fully understand the impact of risk adjustment. The Committee also cautioned about the use of race as a variable and stressed it should not be used as a proxy for SES.
    - Note: These additional analyses were added to the measure submission on March 31, 2017.
- The developer provided additional details on their analyses supporting their decision not to include dual eligible status in the risk adjustment model. The developer ran analyses using two versions of a risk adjustment model that account for dual status by including either (i) separate flags for full and partial dual status or (ii) one flag for full or partial dual status. The first analysis shows the distribution of the difference between a hospital's MSPB-Hospital measure score when calculating expected cost normally and when calculating expected cost using one of the two models accounting for dual status mentioned above. The difference was calculated as (MSPB-Hospital score) – (MSPB-Hospital score, adjusted for dual enrollment). The developer also calculated the mean MSPB-Hospital measure score for providers based on the percent of beneficiaries with full or partial dual status. Finally, the developer conducted an analysis of episodes for non-dual beneficiaries and episodes for dual beneficiaries to identify whether the measure score for hospitals that treat high proportions of dual beneficiaries are affected significantly by their dual population's episodes or not.

- The developer found that the standard deviation of the difference between provider's MSPB-Hospital measure score when calculated normally and when risk adjusting for dual beneficiaries is 0.003. The developer noted that the standard deviation did not change based on using separate flags for full and partial dual status or when using a single flag for full or partial dual status.
- The developer also provided analyses showing distribution of the difference in MSPB-Hospital measure scores when adjusting for beneficiary dual status. When using separate flags for full dual status and partial dual status, the 0.1th percentile has a difference of -0.025, while the 99.9th percentile has a difference of 0.020. When using a single flag for full dual status or partial dual status, the 0.1th percentile has a difference of -0.029, while the 99.9th percentile has a difference of 0.017.
- The developer also examined the mean MSPB-Hospital measure score for hospitals, broken out by percent of beneficiaries with any episodes with full or partial dual status. The developer did not identify a clear trend for MSPB-Hospital measure scores based on the percent of dual-eligible beneficiaries, although hospitals with greater than 60% of beneficiaries having a dual-eligible episode have a higher MSPB-Hospital score on average. The developer also provided the mean MSPB-Hospital measure score when restricting a hospital's episodes to their dual and non-dual episodes respectively. These scores are similar, with scores for non-dual episodes being slightly higher for hospitals with 20 percent or higher of their beneficiaries having a dual episode.
- The developer interpreted these findings as demonstrating that there is a low impact of including the dual-eligibility flag as an social risk factor in risk adjustment. The developer stated that most providers had a very minor change in their MSPB-Hospital measure score and that the tails of the distributions are not disproportionately affected, as the overall magnitude of the change is low for almost all hospitals. The developer also found that specific hospitals are not affected by the inclusion of a dual enrollment flag, since measure scores do not vary much by percent of population that is dual status and measure scores are stable within a hospital across the dual and non-dual beneficiary populations. The developer concluded that including beneficiary dual status in the risk adjustment model has a minimal impact on MSPB-Hospital measure score.

---

3. Feasibility: H-12; M-5; L-0; I-0

(3a. Data is readily available; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The developer indicates that all data elements are in defined fields in electronic claims.
- The developer states the measure's risk adjustment model utilizes the new version of the CMS-HCC methodology, which accounts for the conversion to ICD-10 codes.
- The measure is already in use. During 30-day preview periods, neither the developer nor CMS received reports about measure errors from the measured hospitals (i.e., IPPS hospitals with at least 25 episodes in the performance period).
- Overall, the Committee found the measure feasible and agreed with the developer's assertion that the data elements (i.e., administrative claims) are routinely generated and do not cause increased demands on practitioners.

- In the Committee's pre-evaluation survey, one response noted that while the measure is feasible for entities like the Centers for Medicare and Medicaid Services, it would be difficult for other smaller entities to calculate the measure independently.

---

4. Usability and Use: H-5; M-10; L-3; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The measure is currently used in the Hospital Inpatient Quality Report (IQR) Program and Hospital Value-Based Purchasing (HVBP) Program and available on the Hospital Compare website. The Committee agreed its usage in these programs demonstrates the measure's high level of usability and use.
- Committee members raised concerns that the reports provided on this measure may not be fully actionable, as the information provided does not provide adequate details to show where improvement efforts should be focused. The Committee suggested the measure's usability could be enhanced by providing a more detailed breakdown of utilization by major diagnostic categories in the measure summary reports that are sent to providers.

---

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-17**; **N-1**

- The Committee agreed the measure met the NQF criteria, and voted to recommend it for continued endorsement.

---

6. Public and Member Comment: April 20, 2017 – May 19, 2017

Comments Received

- This measure received five comments, including one supporting its endorsement. Two commenters noted that this measure is only validated and endorsed at the facility level. Commenters expressed concerns that this measure should be validated at the clinician level before adoption for the Merit-Based Incentive Payment System (MIPS) and urged the developer to expedite clinician level testing. Three comments address concerns with the measure's testing for reliability and validity. Commenters questioned the weak association between this measure and measures of readmission. Commenters also raised concerns that the majority of variation in the measure is drive by post-acute spending. Commenters noted this measure is used in the Hospital Value-Based Purchasing Program and that there is a potential for negative unintended consequences from its use. Commenters were also concerned about the SDS adjustment. One comment was received on the general draft report requesting clarification on how all three measures address cancer patients. The commenter noted that there can be significant variation in treatment needs, comorbidities, and patient preferences that can influence cost and resource use.

Developer Response

Response 1: SDS Adjustment

The measure developer appreciates the AMA's feedback on the MSPB-Hospital measure construction and the testing of sociodemographic (SDS) factors in the measure's risk adjustment model. The developer believes that the MSPB-Hospital measure does meet the scientific acceptability criteria of validity, and the NQF committee agreed that the measure met the Scientific Acceptability criterion. The NQF committee had 4 members vote that the measure demonstrated high validity, 9 members vote that the measure demonstrated medium validity, and 5 members vote that the measure had low reliability.

The MSPB-Hospital measure aims to improve care coordination in the period between 3 days prior to an acute inpatient hospital admission through the period 30 days after discharge. The MSPB-Hospital measure recognizes lower costs associated with a reduction in unnecessary services, preventable complications, readmissions, and shifting post-acute care from more expensive to less expensive services when appropriate. The MSPB-Hospital measure creates parallel incentives for hospitals and post-acute care providers. The developer would also like to clarify that 84% of the variance in episode cost is accounted for by post-acute care costs, rather than 84% of total episode costs being attributed to the hospital during the 30 day post-discharge period. This finding is consistent with expectations. The risk adjustment model predicts a certain level of post-discharge spending based upon the beneficiary's prior health history and MS-DRG. Specifically, the MSPB-Hospital risk adjustment methodology adjusts the MSPB-Hospital measure for age, severity of illness, and enrollment status indicators.

Variance in provider scores based on post-discharge spending emphasizes the importance of care transitions and care coordination in improving patient care. Hospitals receive a Hospital-Specific Report (HSR) that provides information on the hospital's performance on the MSPB-Hospital measure, as well as three supplementary hospital-specific data files (an index admission file, a beneficiary risk score file, and an MSPB-Hospital episode file) related to the hospital's MSPB-Hospital measure. Together, these files provide an overview of how the hospital performed on the MSPB-Hospital measure and identify other providers involved in care for their beneficiaries, which facilitates better coordination of care with those providers. No evidence of unintended consequences to individuals or populations, such as changes in referral patterns, have been identified during testing and since implementation.

The developer would also like to note that they submitted an updated measure testing form to the NQF on March 31st, 2017 that contained an appendix with additional analyses responding to NQF feedback and further description of the original submission. That appendix notes that analyses comparing the MSPB-Hospital measure with the condition-specific readmission measures were excluded in the 2016 submission because the condition-specific readmission measures examine hospital performance on a specific set of conditions, while the MSPB-Hospital measure is intended to capture hospital performance across all acute conditions. Consequently, comparisons could be misleading. Since MSPB-Hospital is an all cost measure that includes all conditions, the developer believes that it is more appropriate to look at the correlation between MSPB-Hospital and another broad-based all cost measure (i.e., the risk-adjusted, standardized total Medicare spending at the Hospital Referral Regions (HRR) level). The developer agrees that the MSPB-Hospital measure is most meaningful when presented in the context of other quality measures, which are part of the Hospital Value-Based Purchasing (VBP) Program. As part of the Hospital VBP Program, the MSPB-Hospital measure is combined with current quality of care measures to facilitate profiling hospital value (payments and quality).

The developer believes that the MSPB-Hospital measure submission did meet the requirements of the NQF's SDS trial period and the NQF committee confirmed this by passing the MSPB-

Hospital measure on the Scientific Acceptability criterion. The developer noted in the original submission that the inclusion of SDS factors (i.e., family income-to-poverty ratio and race) had a minimal impact on hospital's measure scores. The developer recognizes the commenter's concerns that additional factors could be included in the SDS measure testing. The developer selected family income-to-poverty ratio to strike a balance between the individual and community factors related to SES and listed by the commenter, as individual family members may pool financial resources to provide care for older relatives. The developer also conducted additional analyses based on feedback from the NQF committee to examine the impact of including a dual eligibility flag in the risk adjustment model, which are included in the appendix of the measure testing form that was submitted to the NQF on March 31st, 2017. These analyses showed that including a dual eligibility flag had a low impact on MSPB-Hospital measure scores and that hospitals on the tails of score distributions were not disproportionately affected. A recent ASPE report showed some differences in measure performance between hospitals with a high amount of Disproportionate Share Hospital payments and a low amount.* The analysis in the appendix's Supplementary Table 7 suggests that these differences may be driven by hospitals with a very high concentration of dual eligible beneficiaries (above 60%), and that measure scores are high for both duals and non-duals in these hospitals. This suggests that these hospitals are relatively higher-cost hospitals for all types of patients.

The MSPB-Hospital measure developer appreciates the commenter's feedback on the separate clinician-level measure (MSPB-TIN) used in the Merit-Based Incentive Payment System (MIPS). The developer would like to clarify that while MSPB-TIN and the facility-level MSPB-Hospital measure currently under consideration for NQF re-endorsement are alike, MSPB-TIN differs in attribution methodology. The MSPB-TIN measure is still under reevaluation. To ensure the reliability and validity of the measures being implemented, CMS reevaluates the measures annually and plans NQF submission of the measures by taking into account program needs and measure implementation timelines to meet the statutory requirements. MSPB-TIN was finalized for inclusion in the MIPS Cost Category as part of the Quality Payment Program Final Rule. The first performance period for the Cost Category is calendar year 2017, and the category is weighted at zero percent for the associated payment year, meaning that it will not impact payments under the program in its first year.

*Office of the Assistant Secretary for Planning and Evaluation (ASPE). "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs." December, 2016. Available at https://aspe.hhs.gov/system/files/pdf/253971/ASPESESRTCfull.pdf.

Response 2: Measure Testing Concerns

The measure developer appreciates FAH's feedback on the MSPB-Hospital measure construction and the testing of sociodemographic (SDS) factors in the measure's risk adjustment model. The developer believes that the MSPB-Hospital measure does meet the scientific acceptability criteria of validity, and the NQF committee agreed that the measure met the Scientific Acceptability criterion. The NQF committee had 4 members vote that the measure demonstrated high validity, 9 members vote that the measure demonstrated medium validity, and 5 members vote that the measure had low reliability.

The MSPB-Hospital measure aims to improve care coordination in the period between 3 days prior to an acute inpatient hospital admission through the period 30 days after discharge. The MSPB-Hospital measure recognizes lower costs associated with a reduction in unnecessary services, preventable complications, readmissions, and shifting post-acute care from more expensive to less expensive services when appropriate. The MSPB-Hospital measure creates

parallel incentives for hospitals and post-acute care providers. The developer would also like to clarify that 84% of the variance in episode cost is accounted for by post-acute care costs, rather than 84% of total episode costs being attributed to the hospital during the 30 day post-discharge period. This finding is consistent with expectations. The risk adjustment model predicts a certain level of post-discharge spending based upon the beneficiary's prior health history and MS-DRG. Specifically, the MSPB-Hospital risk adjustment methodology adjusts the MSPB-Hospital measure for age, severity of illness, and enrollment status indicators.

Variance in provider scores based on post-discharge spending emphasizes the importance of care transitions and care coordination in improving patient care. Hospitals receive a Hospital-Specific Report (HSR) that provides information on the hospital's performance on the MSPB-Hospital measure, as well as three supplementary hospital-specific data files (an index admission file, a beneficiary risk score file, and an MSPB-Hospital episode file) related to the hospital's MSPB-Hospital measure. Together, these files provide an overview of how the hospital performed on the MSPB-Hospital measure and identify other providers involved in care for their beneficiaries, which facilitates better coordination of care with those providers. No evidence of unintended consequences to individuals or populations, such as changes in referral patterns, have been identified during testing and since implementation.

The developer would also like to note that they submitted an updated measure testing form to the NQF on March 31st, 2017 that contained an appendix with additional analyses responding to NQF feedback and further description of the original submission. That appendix notes that analyses comparing the MSPB-Hospital measure with the condition-specific readmission measures were excluded in the 2016 submission because the condition-specific readmission measures examine hospital performance on a specific set of conditions, while the MSPB-Hospital measure is intended to capture hospital performance across all acute conditions. Consequently, comparisons could be misleading. Since MSPB-Hospital is an all cost measure that includes all conditions, the developer believes that it is more appropriate to look at the correlation between MSPB-Hospital and another broad-based all cost measure (i.e., the risk-adjusted, standardized total Medicare spending at the Hospital Referral Regions (HRR) level). The developer agrees that the MSPB-Hospital measure is most meaningful when presented in the context of other quality measures, which are part of the Hospital Value-Based Purchasing (VBP) Program. As part of the Hospital VBP Program, the MSPB-Hospital measure is combined with current quality of care measures to facilitate profiling hospital value (payments and quality).

The developer believes that the MSPB-Hospital measure submission did meet the requirements of the NQF's SDS trial period and the NQF committee confirmed this by passing the MSPB-Hospital measure on the Scientific Acceptability criterion. The developer noted in the original submission that the inclusion of SDS factors (i.e., family income-to-poverty ratio and race) had a minimal impact on hospital's measure scores. The developer recognizes the commenter's concerns that additional factors could be included in the SDS measure testing. The developer selected family income-to-poverty ratio to strike a balance between the individual and community factors related to SES and listed by the commenter, as individual family members may pool financial resources to provide care for older relatives. The developer also conducted additional analyses based on feedback from the NQF committee to examine the impact of including a dual eligibility flag in the risk adjustment model, which are included in the appendix of the measure testing form that was submitted to the NQF on March 31st, 2017. These analyses showed that including a dual eligibility flag had a low impact on MSPB-Hospital measure scores and that hospitals on the tails of score distributions were not disproportionately

affected. A recent ASPE report showed some differences in measure performance between hospitals with a high amount of Disproportionate Share Hospital payments and a low amount.* The analysis in the appendix's Supplementary Table 7 suggests that these differences may be driven by hospitals with a very high concentration of dual eligible beneficiaries (above 60%), and that measure scores are high for both duals and non-duals in these hospitals. This suggests that these hospitals are relatively higher-cost hospitals for all types of patients.

*Office of the Assistant Secretary for Planning and Evaluation (ASPE). "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs." December, 2016. Available at https://aspe.hhs.gov/system/files/pdf/253971/ASPESESRTCfull.pdf.

Response 3: Clinician-Level Measurement

The MSPB-Hospital measure developer appreciates the commenter's feedback on the separate clinician-level measure (MSPB-TIN) used in the Physician Value Modifier (VM) program and slated for use in the Merit-Based Incentive Payment System (MIPS). The developer would like to clarify that while MSPB-TIN and the facility-level MSPB-Hospital measure (NQF #2158) currently under consideration for NQF re-endorsement are alike, the MSPB-TIN measure for MIPS differs in attribution methodology. The MSPB-Hospital measure (NQF #2158) is not slated for use in the MIPS program, and this NQF submission does not cover the clinician-level MSPB-TIN measure. The MSPB-TIN measure for MIPS is still under reevaluation. To ensure the reliability and validity of the measures being implemented, CMS reevaluates the measures annually and plans NQF submission of the measures by taking into account program needs and measure implementation timelines to meet the statutory requirements. MSPB-TIN was finalized for inclusion in the MIPS Cost Category as part of the Quality Payment Program Final Rule (https://www.federalregister.gov/documents/2016/11/04/2016-25240/medicare-program-merit-based-incentive-payment-system-mips-and-alternative-payment-model-apm). The first performance period for the Cost Category is calendar year 2017, and the category is weighted at zero percent for the associated payment year, meaning that it will not impact payments under the program in its first year.

Response 4: Inclusion of Cancer Patients

Thank you for your comment. We recognize that cancer patients often have complex comorbidities and require more intensive treatment. The MSPB-Hospital measure accounts for comorbidities through risk adjustment. Specifically, the MSPB-Hospital risk adjustment methodology adjusts the MSPB-Hospital measure for age, severity of illness, and enrollment status indicators. The methodology includes 12 age-categorical variables; 79 hierarchical condition category (HCC) variables derived from the beneficiary's claims during the period 90 days prior to the start of the episode to measure severity of illness; as well as the MS-DRG of the index hospitalization. The risk adjustment methodology also includes the HCC interaction variables, status indicator variables for whether the beneficiary qualifies for Medicare through Disability or End-Stage Renal Disease (ESRD), and whether a beneficiary resides in a long-term care facility. The MS-DRG is included in the risk adjustment model to better account for the differences in cost of care that stem from different reasons for hospitalization, including cancer. This allows the MSPB-Hospital measure to compare cost of care across all conditions, rather than focusing on a specific disease. As such, the risk factor of the MS-DRG of the index hospitalization should account for the more intensive treatment that cancer patients may require.

We appreciate your comment that cost measures should capture and categorize costs throughout the cycle of care. Hospitals that have an MSPB-Hospital measure receive a Hospital-Specific Report that provides a cost breakdown by claim type for the hospital's MSPB-Hospital episodes for three categories: 3 days prior to index admission, during-index admission, and 30 days after hospital discharge. This breakdown is provided for informational purposes to allow hospitals to evaluate its episode spending before, during, or after the index hospital admission.

We also wanted to acknowledge and address your comment that cost measures are not necessarily good markers of quality. For this reason, we note in our public documentation that the MSPB-Hospital measure alone is not intended to necessarily reflect the quality of care provided by hospitals. Accordingly, a lower MSPB-Hospital measure score across performance periods (i.e., lower Medicare spending per beneficiary) in isolation, should not be interpreted as better care. The MSPB Measure is most meaningful when presented in the context of other quality measures, which are part of the Hospital Value-Based Purchasing (VBP) Program. As part of the Hospital VBP Program, the MSPB-Hospital measure is combined with current quality of care measures to facilitate profiling hospital value (payments and quality).

Committee Response

The Committee had in-depth conversations on the attribution of #2158. The Committee recognizes that hospitals may not have complete control over the spending captured by the measure. However, the Committee believes that there are actions hospitals can take to improve their performance on this measure. Additionally, the Committee noted the need for attribution models that support care coordination and team-based care as the system aims to transition from fee-for-service to population-based payment.

Consideration of social risk factors in risk adjustment models is a critical issue in measurement science. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer. The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and social risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between social risk factors and cost and resource use.

While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee recognized the impact that social risk can have on cost and resource use measures and encourages measure developers to test the impact of additional social risk variables. The Committee also encouraged exploration of the impact of community-level variables. However, the Committee generally agreed that the risk adjustment method used in these measures met the NQF criteria given the data available to the developer, and the measure testing results presented.

The Committee agrees that the measure is only validated and recommended for use at the facility level, and needs further testing before it can be considered for endorsement at the physician level.

The Committee has reviewed the comments and appreciates the additional insights on the measure. After reviewing the comments and responses from the developer, the Committee believes this measure is appropriately specified and tested and continues to meet the criteria for NQF endorsement.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-14; N-0

- **Decision**: Approved for continued endorsement

8. Appeals

No appeals received.

Appendix B: NQF Cost and Resource Use Portfolio in Federal Programs

NQF #	Title	Federal Program
2431	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode-of-Care for Acute Myocardial Infarction (AMI)	Hospital Inpatient Quality Reporting Program, Hospital Value-Based Purchasing Program
2436	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode-of-Care for Heart Failure (HF)	Hospital Inpatient Quality Reporting Program, Hospital Value-Based Purchasing Program
2579	Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Pneumonia	Hospital Inpatient Quality Reporting Program
2158	Medicare Spending Per Beneficiary (MSPB)	Hospital Inpatient Quality Reporting Program, Hospital Value-Based Purchasing Program

Appendix C: Cost and Resource Use Standing Committee and NQF Staff

STANDING COMMITTEE

Brent Asplin, MD, MPH (Co-Chair)
Independent
Cincinnati, OH

Cheryl Damberg, PhD (Co-Chair)
RAND Corporation
Santa Monica, CA

Larry Becker
Retired
Rochester, NY

Mary Ann Clark, MHA
Avalere
Washington, DC

Jennifer Eames Huff, MPH, CPEH
JEH Consulting
San Francisco, CA

Troy Fiesinger, MD, FAAFP
Village Family Practice of Fort Bend
Houston, TX

Nancy Garrett, PhD
Hennepin County Medical Center
Minneapolis, MN

Andrea Gelzer, MD, MS, FACP
AmeriHealth Caritas
Philadelphia, PA

Lisa Latts, MD, MSPH, MBA, FACP *(Inactive 2016-2017)*
IBM
Denver, CO

Martin Marciniak, MPP, PhD
GlaxoSmithKline
Durham, NC

Kristine Martin Anderson, MBA
Booz Allen Hamilton
Rockville, MD

James Naessens, ScD, MPH
Mayo Clinic
Rochester, MN

Jack Needleman, PhD
UCLA Fielding School of Public Health
Los Angeles, CA

Janis Orlowski, MD, MACP
Association of American Medical Colleges
Washington, DC

Carolyn Pare *(Inactive 2016-2017)*
Minnesota Health Action Group
Bloomington, MN

Betty Rambur, PhD, RN
University of Rhode Island
Kingston, RI

John Ratliff, MD, FACS, FAANS
American Association of Neurological Surgeons
Stanford, CA

Andrew Ryan, PhD *(Inactive 2016-2017)*
Weill Cornell Medical College
New York, NY

Srinivas Sridhara, PhD, MHS
The Advisory Board Company
Washington, DC

Lina Walker, PhD *(Inactive 2016-2017)*
Public Policy Institute
Washington, DC

Bill Weintraub, MD, FACC
Christiana Care Health System
Newark, DE

Herbert Wong, PhD
Agency for Healthcare Research and Quality
Rockville, MD

Dolores Yanagihara, MPH
Integrated Healthcare Association
Oakland, CA

NQF STAFF

Helen Burstin, MD, MPH
Chief Scientific Officer

Marcia Wilson, PhD, MBA
Former Senior Vice President, Quality Measurement

Elisa Munthali, MPH
Acting Senior Vice President, Quality Measurement

Erin O' Rourke
Senior Director

Ashlie Wilbon, MS, MPH, FNP-C
Senior Director

Rachel Roiland, PhD, RN
Former Senior Project Manager

Suzanne Theberge, MPH
Senior Project Manager

Hiral Dudhwala, RN, MSN/MPH
Project Manager

Irvin Singh, MPH
Project Analyst

CONSULTANT

Taroon Amin, PhD, MPH
Project Consultant

Appendix D: Measure Specifications

1598 Total Resource Use Population-Based PMPM Index

STEWARD

HealthPartners

DESCRIPTION

The Resource Use Index (RUI) is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.

A Total Cost of Care Index (NQF-endorsed #1604) when viewed together with the Total Resource Use measure provides a more complete picture of population based drivers of health care costs.

TYPE

Maintenance

DATA SOURCE

Claims (Only)

LEVEL

Clinician: Group/Practice, Population: Community, County or City

SETTING

All care settings and conditions

NUMERATOR STATEMENT

Total Resource PMPM

NUMERATOR DETAILS

(Total Medical TCRRV / Medical Member Months) + (Total Pharmacy TCRRV / Pharmacy Member Months)

DENOMINATOR STATEMENT

Average Risk Score

DENOMINATOR DETAILS

The medical claims data is submitted through the Johns Hopkins ACG Risk Grouper which generates a relative risk score for each member. That risk score is then multiplied by the number of months a member has been enrolled creating a risk weight. The risk weights are then summed to the desired level of measurement (e.g., provider group) and divided by the total sum of the desired level's member months creating a member month weighted Average Risk Score.

ACG Adjusted Total Resource Use PMPM = Total Resource Use PMPM / ACG Risk Score

Resource Use Index = Provider ACG Adjusted Total Resource Use PMPM / Peer Group ACG Adjusted Total Resource Use PMPM

EXCLUSIONS

- Members over age 64
- Members under age 1
- Member enrollment less than nine months during the one year measurement time window
- Members who are not attributed to a primary care provider
- Dollars per member above $125,000

EXCLUSION DETAILS

The HealthPartners' Total Resource Use measure is a full population-based measure, with members under age 1, members 65+ and members with less than 9 months of enrollment excluded to ensure an accurate risk assessment is made on the population.

RISK ADJUSTMENT

- The measure is risk adjusted for age, gender, and diagnosis using the Adjusted Clinical Group (ACG) method.
- The ACG method involves:
    - Grouping International Classification Diagnosis (ICD) diagnosis codes into 32 diagnosis groups (i.e., Aggregated Diagnosis Groups (ADGs)). These ADGs are clinically similar and expected to have similar need for healthcare resources.
    - Adjusted Clinical Groups (ACGs) are created from the ADG assignments and are defined by morbidity, age, and sex. Individual members are then assigned to a single ACG category, which quantifies their risk.

STRATIFICATION

Measures are adjusted for clinical risk and limited to the commercial population.

TYPE SCORE

Ratio

ALGORITHM

Measure was tested using commonly used Attribution Algorithm in an open access market (plurality model, using most recent visit as a tie breaker):

- Include twelve months based on first date of service for the measurement year (e.g. January 1 – December 31) of professional claims experience, with three months of paid claims run out to allow for claims lag.
- Exclude all services that are not office based
- Exclude convenience care clinic visits and hospice services
- Exclude a providers that are not a physician, physician assistant or nurse practitioner
- Assign each service line a specialty based on the servicing physician's practicing specialty or credential specialty if practicing

  Specialty is not available.

- Include only the following specialties: Family Medicine, Internal Medicine, Pediatrics, Geriatrics, OB/GYN

COPYRIGHT / DISCLAIMER

1604 Total Cost of Care Population-Based PMPM Index

STEWARD

HealthPartners

DESCRIPTION

Total Cost of Care reflects a mix of complicated factors such as patient illness burden, service utilization and negotiated prices. Total Cost Index (TCI) is a measure of a primary care provider's risk adjusted cost effectiveness at managing the population they care for. TCI includes all costs associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.

A Total Cost of Care Index when viewed together with HealthPartners (NQF-endorsed #1598) Total Resource Use measure provides a more complete picture of population based drivers of health care costs.

TYPE

Per capita (population- or patient-based)

DATA SOURCE

Claims (Only)

LEVEL

Clinician: Group/Practice, Population : Community, County or City

SETTING

This is a population-based measure that applies to all care settings and conditions.

NUMERATOR STATEMENT

Total PMPM

NUMERATOR DETAILS

(Total Medical Cost / Medical Member Months) + (Total Pharmacy Cost / Pharmacy Member Months)

DENOMINATOR STATEMENT

Average Risk Score

DENOMINATOR DETAILS

The medical claims data is submitted through the Johns Hopkins ACG Risk Grouper which generates a relative risk score for each member. That risk score is then multiplied by the number of months a member has been enrolled creating a risk weight. The risk weights are then summed to the desired level of measurement (e.g., provider group) and divided by the total sum of the desired level's member months creating a member month weighted Average Risk Score.

ACG Adjusted PMPM = Total PMPM / ACG Risk Score

TCI = Provider ACG Adjusted PMPM / Peer Group ACG Adjusted PMPM

EXCLUSIONS

- Members over age 64
- Members under age 1
- Member enrollment less than nine months during the one year measurement time window
- Dollars per member up to $125,000 are included; dollars per member above $125,000 are excluded (truncated)

EXCLUSION DETAILS

The HealthPartners' Total Cost of Care measure is a full population-based measure, with members under age 1, members 65+ and members with less than 9 months of enrollment excluded to ensure an accurate risk assessment is made on the population.

RISK ADJUSTMENT

- The measure is risk adjusted for age, gender, and diagnosis using the Adjusted Clinical Group (ACG) method.
- The ACG method involves:
    - Grouping International Classification Diagnosis (ICD) diagnosis codes into 32 diagnosis groups (i.e., Aggregated Diagnosis Groups (ADGs)). These ADGs are clinically similar and expected to have similar need for healthcare resources.
    - Adjusted Clinical Groups (ACGs) are created from the ADG assignments and are defined by morbidity, age, and sex. Individual members are then assigned to a single ACG category, which quantifies their risk.

STRATIFICATION

Measures are adjusted for clinical risk and limited to the commercial population.

TYPE SCORE

Ratio

ALGORITHM

Measure was tested using commonly used Attribution Algorithm in an open access market (plurality model, using most recent visit as a tie breaker):

- Include twelve months based on first date of service for the measurement year (e.g. January 1 – December 31) of professional claims experience, with three months of paid claims run out to allow for claims lag.
- Exclude all services that are not office based
- Exclude convenience care clinic visits and hospice services
- Exclude a providers that are not a physician, physician assistant or nurse practitioner
- Assign each service line a specialty based on the servicing physician's practicing specialty or credential specialty if practicing

  Specialty is not available.
- Include only the following specialties: Family Medicine, Internal Medicine, Pediatrics, Geriatrics, OB/GYN

COPYRIGHT / DISCLAIMER

2158 Medicare Spending Per Beneficiary (MSPB) – Hospital

STEWARD

Centers for Medicare and Medicaid Services

DESCRIPTION

The Medicare Spending Per Beneficiary (MSPB) - Hospital measure evaluates hospitals' risk-adjusted episode costs relative to the risk-adjusted episode costs of the national median hospital. Specifically, the MSPB-Hospital measure assesses the cost to Medicare for services performed by hospitals and other healthcare providers during an MSPB-Hospital episode, which is comprised of the periods immediately prior to, during, and following a patient's hospital stay. The MSPB-Hospital measure is not condition specific and uses standardized prices when measuring costs. Beneficiary populations eligible for the MSPB-Hospital calculation include Medicare beneficiaries enrolled in Medicare Parts A and B who were discharged from short-term acute Inpatient Prospective Payment System (IPPS) hospitals during the period of performance.

TYPE

Cost per episode

DATA SOURCE

Claims (Only); Other

LEVEL

Facility

SETTING

Hospital: Acute Care Facility

NUMERATOR STATEMENT

Hospital's MSPB-Hospital amount

NUMERATOR DETAILS

Average spending level for the hospital's MSPB-Hospital episodes divided by the average expected episode spending level for the hospital's episodes, multiplied by the average spending over all episodes across all hospitals nationally.

DENOMINATOR STATEMENT

The episode-weighted median MSPB-Hospital amount across all episodes nationally.

DENOMINATOR DETAILS

This distribution of hospitals' MSPB Measure values is provided to hospitals as part of their hospital-specific reports (HSRs). The denominator of the MSPB-Hospital measure is weighted by the number of episodes; as a result, the (unweighted) median MSPB-Hospital measure score is not necessarily always equal to one.

EXCLUSIONS

- Acute-to-acute transfer episodes: based on claim discharge code
- Death episodes: beneficiary dies during the measurement episode
- Overlapping episodes: occurrence of an inpatient admission during the 30 days post-discharge of an index admission is not considered a new index admission
- Outlier episodes: episode whose relative scores fall above the 99th percentile or below the 1st percentile of the distribution of residuals

EXCLUSION DETAILS

Acumen evaluated the validity of the measure exclusion criteria by producing impact analyses, which show the effect of recalculating the MSPB-Hospital measure while independently reversing each of the following exclusion criteria: (1) acute-to-acute transfer episodes; (2) death episodes; and (3) outlier episodes. For (1), our analysis evaluated the impact of including transfer episodes on MSPB-Hospital measure scores. For (2), we re-calculated the MSPB-Hospital measure using beneficiaries who die during the episode. Specifically, we examined the percent of beneficiaries who die during the MSPB-Hospital episode and the effect that including death episodes had on hospital scores. For (3), we examined the effect of including outliers when calculating MSPB-Hospital measure scores instead of excluding outliers based on the distribution of residuals. Specifically, we examined the impact of top-coding episodes with risk-adjusted costs that are above the 99th percentile, where those episodes are assigned the cost of the episode at the 99th percentile. We also examined the impact of bottom-coding episodes with risk-adjusted costs that are below the 1st percentile, where those episodes are assigned the cost of the episode at the 1st percentile.

The measure also implements an exclusion criteria specific to inpatient admissions that are allowed to trigger a new MSPB-Hospital measure. Specifically, we do not allow inpatient admissions that occur within 30 days post-discharge of another inpatient admission to start a new MSPB-Hospital episode; we refer to this criteria as excluding overlapping episodes. For this exclusion (4), we analyzed the effect of including overlapping episodes when constructing the MSPB-Hospital episodes. To illustrate what this exclusion is, take an inpatient admission that triggers Episode A and see if the beneficiary has another inpatient admission within the 30-day post-discharge window of Episode A. If the beneficiary has a second qualifying admission within the 30-day post-discharge window of Episode A, do not allow the second admission to trigger Episode B. We evaluated the impact of this exclusion on MSPB-Hospital measures by re-calculating MSPB-Hospital with the previously-excluded episodes added back in, which was then compared to MSPB-Hospital measures calculated under the overlapping episodes exclusion.

The developer found the following:

1. *Transfer Episodes:* Episodes that include an acute-to-acute transfer account for 1.6% of total episodes. Episodes containing an acute-to-acute transfer have an average observed cost of $33,363 compared to an average expected cost of $21,068, resulting in an observed-to-expected cost ratio of 1.58. Episodes not containing an acute-to-acute transfer, on the other hand, have an average observed cost of $20,570 compared to an average expected cost of $20,774, resulting in a observed-to-expected cost ratio of 0.99. Rural hospitals tend to have a higher rate of transfers than urban hospitals (4.1% and 1.3%, respectively), so including transfer episodes that have higher observed-to-expected cost ratio in the MSPB-Hospital measure calculation would probably disproportionately worsen rural hospitals' scores. When including transfer episodes in the calculation of the MSPB-Hospital measure, 81% of hospitals' MSPB-Hospital measure scores change by less than ±0.03, and less than 2% of hospitals' MSPB-Hospital measure scores change by more than ±0.10. The correlation between MSPB-Hospital measure scores when excluding transfer episodes versus when including transfer episodes is 0.95.

2. *Death Episodes:* In approximately 8% of MSPB-Hospital episodes, the beneficiary dies before the end of the 30-day post-discharge period. Episodes in which the beneficiary dies during the episode window (denoted as "death episodes") appear more efficient than non-death episodes. The average observed cost of death episodes is $21,041 compared to the expected cost of $24,980, resulting in an observed-to-expected cost ratio of 0.84. Comparatively, non-death episodes have an observed-to-expected cost ratio of 1.02 ($20,512 over $20,156). If death is included in measure calculation, 96% of hospitals' MSPB-Hospital measure scores change by less than ±0.03, and very few hospitals (less than 0.2%) see changes in MSPB-Hospital measure scores greater than ±0.10. The correlation between MSPB-Hospital measure scores when excluding death episodes versus when allowing for inclusion of death episodes in measure calculation is 0.99.

3. *Outlier Episodes:* When including outlier episodes in measure calculation, about 2% of hospitals see an absolute change in their MSPB-Hospital measure score of greater than ±0.10, and 6% of hospitals' MSPB-Hospital measure scores change by greater than ±0.05. The correlation between MSPB-Hospital measure scores when excluding outliers versus when including outliers is 0.93.

4. *Overlapping Episodes:* Approximately 12% of episodes had their trigger inpatient admission within 30 days of the discharge date of the trigger inpatient admission of another episode. If episodes with a trigger inpatient admission during the 30-day post-discharge period of another episode are included in MSPB-Hospital measure calculation, 97% of hospitals' MSPB-Hospital measure scores change by less than ±0.03, with a small proportion of hospitals (0.4%) experiencing changes in MSPB-Hospital measure scores greater than ±0.10. The correlation of MSPB-Hospital measure scores before and after removing the overlapping episodes exclusion is 0.99.

The developer interpreted these findings as follows:

1. *Transfer Episodes:* Because transfer episodes are more inefficient than non-transfer episodes, regardless of the type of hospital (urban or rural), there are two main problems with including transfer episodes. First, because the observed cost relative to the predicted cost is high for transfer episodes (partly due to partial or full payments for two inpatient stays), including transfer episodes in the MSPB-Hospital measure may likely increase the MSPB-Hospital measure score of those hospitals most often engaging in transfers. These hospitals may not always have the capacity to handle these cases, and CMS may have an interest in ensuring medically appropriate transfers occur. Second, excluding transfer episodes addresses stakeholder concerns that neither the admitting nor receiving hospital is fully able to coordinate care. Stakeholders find it inappropriate to hold the transferring hospital responsible for services rendered by the receiving hospital, and it also may not be appropriate to hold the receiving hospital responsible for issues that arose prior to admission of a transferred patient. As a result, transfer episodes are excluded from the MSPB-Hospital measure calculation.

2. *Death Episodes:* Cases where the beneficiary dies during the episode are not eligible to be included in the MSPB-Hospital measure. Though the difference between cost for death and non-death episodes is relatively small compared to other exclusions, there are a few explanations for the exclusion of death episodes. First, including death episodes in MSPB-Hospital measure calculation may create problematic incentives. Death episodes appear more efficient than non-death episodes; unlike non-death episodes, which have a slightly greater observed cost than expected cost, the observed cost for death episodes is much less than the expected cost. This is because beneficiaries with death episodes likely have shorter episodes (and therefore fewer services) than beneficiaries with non-death episodes with the same DRG. Because of this, including death episodes in MSPB-Hospital measure calculation may incentive low-quality care, as increased mortality rates could potentially improve hospitals' MSPB-Hospital measure scores by including episodes that appear more efficient. Second, episodes during which a beneficiary dies are "truncated;" in other words, costs that might have occurred if the beneficiary had not died are not observed due to death. Death episodes are incomplete episodes where significant data could be missing when death occurs early in the episode. To avoid including episodes of care with incomplete costs and problematic incentives, episodes during which a beneficiary dies are excluded from the MSPB-Hospital measure calculation.

3. *Outlier Episodes:* Outliers are excluded from the MSPB-Hospital measure calculation to avoid cases where a handful of high-cost and low-cost outliers have a disproportionate effect on each hospital's MSPB-Hospital measure score. While the correlation between the measure

when excluding outliers versus when including outliers is extremely high (0.93), outlier episodes impact a small percentage of hospitals' MSPB-Hospital measure scores in a large and important way. The distribution of hospital risk-adjusted episode spending is significantly right-skewed: the 99th percentile is 3.6 times the value of the median, while the 1st percentile is less than half the value of the median. Excluding outliers based on risk-adjusted cost eliminates the episodes that deviate most from the spending levels one would have expected based on patient demographics and severity of illness.

4. *Overlapping Episodes*: Episodes that begin during a prior episode's 30-day post-discharge period are excluded from MSPB-Hospital measure calculation. The impact of the exclusion on hospitals' MSPB-Hospital measure scores is minimal, and the correlation of the MSPB-Hospital measure calculated with and without implementing the overlapping episodes exclusion is high.

RISK ADJUSTMENT

Ordinary least squares (OLS) linear regression model based on the Centers for Medicare & Medicaid Services' hierarchical conditions categories (CMS-HCC) risk adjustment methodology. Independent variables included in the model: beneficiary age, health status (measured by hierarchical condition categories (HCCs), disability status, end-stage renal disease (ESRD) status, resident in a long-term care facility, MS-DRG indicators for the index admission, and disease interactions (HCCs x enrollment status).

STRATIFICATION

While the measure results are not stratified, expected costs for episodes are determined by using a separate risk adjustment model for episodes within each MDC. MDCs are aggregations of Diagnosis Related Groups (MS-DRG), which CMS uses to classify acute inpatient admissions.

TYPE SCORE

Ratio

ALGORITHM

The MSPB-Hospital measure assesses the standardized allowed amounts of services performed by hospitals and other healthcare providers during an MSPB-Hospital episode, which includes all Part A and Part B Medicare claims that occur within the time period 3 days prior to the index hospital admission through 30 days after discharge from the index admission. As a result, costs from all Part A and Part B claim types (i.e., inpatient, outpatient, home health agency, hospice, skilled nursing facility, durable medical equipment, and carrier) are included. Note that costs of Part B drugs are included but costs of Part D drugs are not included since Part D is not used to calculate the MSPB-Hospital measure. The methodology used to payment standardize these claims is available for download ("CMS Price (Payment) Standardization").

To assist providers in examining their spending, CMS provides MSPB-Hospital spending breakdowns by different claim types (i.e., home health agency, hospice, inpatient, outpatient, skilled nursing facility, durable medical equipment, and physician/carrier), as well as by time period (i.e., 3 days prior to index admission, during-index admission, and 30 days after hospital discharge).

COPYRIGHT / DISCLAIMER

N/A

Appendix E: Pre-Evaluation Comments

Comments received as of March 06, 2017.

1598 Total Resource Use Population-Based PMPM Index

Submitted by American Medical Association

Given measure 1598 and 1604 are maintenance measures, the AMA would have expected the developer, HealthPartners to have provided more information on actual performance data and how well the measures performed in the real world across different groups. The developer references all of the groups that started collecting the measure as an indicator that there is progress toward improvement, but uptake of a measure does not mean the same thing as improving performance. We, therefore, have the following concerns:

The measure submission documents state that many groups and institutions are collecting and reporting the measure under the testing and usability section, but we are only provided data from HealthPartner groups in Minnesota and Western Wisconsin. We would like for data from the first submission and anything within the last 4 years to be included and for the data to include mean, std dev, min, max, interquartile range, and scores by decile. It is also not clear to us how HealthPartners standardizes prices.

We also seek clarification on the sample size. The document states it has been tested with a minimum attributed population of 600 members, but it is not clear whether this is with each practice group or by payer or plan. The reliability testing discussion also fails to address the sample size question and the number of physicians or patient that must be attributed to a group for the measure to be considered reliable. This issue was raised as a concern when the measure underwent its last review and once, again, we request more clarity around the level of analysis and how a physician group is defined.

We also find the risk-adjustment strategy utilized for this measure insufficient. The developer utilizes the ACG system which is proprietary and groups must pay to use it. The developer states you can use others but no testing of other risk-adjustment strategies is outlined to compare the results of different tools. It would be helpful to know whether the groups that implemented the measure are all using the ACG system. If not, then it is not quite clear whether the measure produces comparable results across institutions. With the SES analysis, we do not believe the developer provided an adequate conceptual analysis or sufficient information on why they did not test one of the two factors. They first state that they looked at two factors (income and education), cite one or two articles and then they say they could only look at one- income. Therefore, we do not believe what was provided is sufficient to satisfy the SES trial requirements.

We also are concerned with the definition of primary care physician because it includes specialties such as OB/GYN that have higher intensity of services. It would also be helpful to have validity testing that includes comparisons across the different specialties that are defined as primary care physicians by the measure developer and then against all of the groups to see if it can distinguish meaningful differences and not yield inaccurate comparisons by specialty.

1598 Total Resource Use Population-Based PMPM Index

Submitted by SelectHealth

SelectHealth supports endorsement of the HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598). These measures have been widely adopted by many stakeholders in the health care community and have advanced the national conversation of health care affordability.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Sanne Jones Magnan

Thank you for the opportunity to share my support for the HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. With my internal medicine background and my previous leadership roles as the Minnesota Commissioner of Health and President & CEO of the Institute for Clinical Systems Improvement, I know firsthand the importance of the Triple Aim for our communities and our patients. The Total Cost of Care and Total Resource Use measures help leaders, decision-makers, and physicians identify improvement opportunities for affordability and value in our healthcare systems. The measures provide transparent information needed to drive change for better health and experience at a lower cost for our patients and communities.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Independent Health

Independent Health supports endorsement of HealthPartner's Total Cost of Care (#1604 and #1598) measures. These measures have been adopted by many stakeholders in the health care community and have advanced the national discussion on health care affordability.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Greenville Hospital System

Greenville Health System fully supports endorsement of the Health Partners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. These measures have been widely adopted by many stakeholders in the healthcare community and have advanced the national conversation around healthcare affordability.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Tufts Health Plan

Tufts Health Plan supports endorsement of the Health Partners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. These measures have been widely adopted by many stakeholders in the health care community and have advanced the national conversation of health care affordability.

1598 Total Resource Use Population-Based PMPM Index

Submitted by HealthPartners Medical Group

HealthPartners Medical Group strongly recommends for endorsement both the Total Cost of Care (#1604) and Total Resource Use (#1598) measures. For more than a decade, Total Cost of Care (TCOC) has been the top line measure of affordability for our care group. We drill down from the overall measure of TCOC to price drivers, and Total Resource Use drivers to identify opportunities for improvement. These measures have guided our improvement strategies; allowing us to focus on appropriate use of services and place of service opportunities. This has resulted in improved affordability for our patients. Our full statement of support and usability of these measures was included in the measure submission.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Andrew Dorwart, MD

Stillwater Medical Group and Lakeview Hospital is an integrated, non-profit clinic and hospital system serving the eastern Twin Cities metro area and Western Wisconsin. We use HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures in our system to identify opportunities to improve affordability for our patients. We support maintaining endorsement of the HealthPartners measures.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Park Nicollet Health Services

Park Nicollet appreciates the opportunity to voice our support for HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. HealthPartners has transparently shared the measurement method and measure results with providers in our community for nearly a decade, and we have used these measures to improve health care affordability for our patients, while maintaining top quality performance. Our full statement of support and comment on the usability and usefulness of these measures was submitted as part of HealthPartners Total Cost of Care and Total Resource Use NQF submission.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Sanford Health

Sanford Health supports endorsement of the HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. As an integrated health system in the HealthPartners network, we appreciate the transparency and soundness of the measures, as well as our partnership with HealthPartners as we strive to improve care for our patients. The Sanford Health Plan is also a licensee and user of the measures.

1598 Total Resource Use Population-Based PMPM Index

Submitted by Network for Regional Healthcare Improvement

On behalf of NRHI, we are in support of NQF endorsing this measure. For over three years we have been actively engaged with regions across the country measuring, reporting and using the total resource use population based PMPM index. Recently we published a benchmark report that utilized this measure and compared across 5 regions which has resulted in meaningful conversations within regions about the cause of variation. Seven regions have produced and distributed attributed practice level reports in their communities at least once, some multiple times over the past few years. During 2015, healthcare cost information on over 5 million patients attributed to 20,000 individual physicians were included in practice level reports and used by practices to identify areas of variation and opportunities for intervention to improve care while decreasing costs. The utility of this measure increases as you are able to isolate resource use - which is very powerful and something physicians can control.

The basic foundation for all of these efforts is the HealthPartners NQF endorsed TCOC measure framework. NRHI has been awarded funding from RWJF for a third phase which began on November 1, 2016. During this two-year grant, we will expand the number of regions producing, sharing and using TCOC for both commercial and Medicare populations, maintain and grow our Getting to Affordability Learning Modules and community - a place to connect with others across the country who are measuring and using TCOC, convene a multi-stakeholder summit on using TCOC to advance the Triple Aim and payment reform, and develop and implement sustainability plans to ensure future ability to produce, share and use TCOC.

We support further endorsement of this measure and would be happy to answer any questions.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by American Hospital Association

The American Hospital Association (AHA) recognizes the importance of total cost of care and resource use measures in helping those running health plans better understand and address opportunities to improve the value of the care provided. Therefore, we are exploring a partnership with HealthPartners to pilot use of their measures (#1604 Total Cost of Care and #1598 Total Resource Use), with the goal of using these measures with a subset of our members with health plans to help them better understand their performance. We look forward to working with HealthPartners on designing and implementing this important pilot to enhance value of care for the patients and communities our member organizations serve. Our full letter was included with the HealthPartners submission documents.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by UPMC Health Plan

UPMC Health Plan supports endorsement of the HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. These measures have been widely adopted by many stakeholders in the health care community and have advanced the national conversation of health care affordability.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by American Medical Association

Given measure 1598 and 1604 are maintenance measures, we would have expected the developer, HealthPartners to have provided more information on actual performance data and how well the measures performed in the real world across different groups. The developer references all of the groups that started collecting the measure as an indicator that there is progress toward improvement, but uptake of a measure does not mean the same thing as improving performance. We, therefore, have the following concerns:

The measure submission documents state that many groups and institutions are collecting and reporting the measure under the testing and usability section, but we are only provided data from HealthPartner groups in Minnesota and Western Wisconsin. We would like for data from the first submission and anything within the last 4 years to be included and for the data to include mean, std dev, min, max, interquartile range, and scores by decile. It is also not clear to us how HealthPartners standardizes prices.

We also seek clarification on the sample size. The document states it has been tested with a minimum attributed population of 600 members, but it is not clear whether this is with each practice group or by payer or plan. The reliability testing discussion also fails to address the sample size question and the number of physicians or patient that must be attributed to a group for the measure to be considered reliable. This issue was raised as a concern when the measure underwent its last review and once, again, we request more clarity around the level of analysis and how a physician group is defined.

We also find the risk-adjustment strategy utilized for this measure insufficient. The developer utilizes the ACG system which is proprietary and groups must pay to use it. The developer states you can use others but no testing of other risk-adjustment strategies is outlined to compare the results of different tools. It would be helpful to know whether the groups that implemented the measure are all using the ACG system. If not, then it is not quite clear whether the measure produces comparable results across institutions. With the SES analysis, we do not believe the developer provided an adequate conceptual analysis or sufficient information on why they did not test one of the two factors. They first state that they looked at two factors (income and education), cite one or two articles and then they say they could only look at one- income. Therefore, we do not believe what was provided is sufficient to satisfy the SES trial requirements.

We also are concerned with the definition of primary care physician because it includes specialties such as OB/GYN that have higher intensity of services. It would also be helpful to have validity testing that includes comparisons across the different specialties that are defined as primary care physicians by the measure developer and then against all of the groups to see if it can distinguish meaningful differences and not yield inaccurate comparisons by specialty.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by Sanne Jones Magnan

Thank you for the opportunity to share my support for the HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. With my internal medicine background and my previous leadership roles as the Minnesota Commissioner of Health and President & CEO of the Institute for Clinical Systems Improvement, I know firsthand the importance of the Triple Aim for our communities and our patients. The Total Cost of Care and Total Resource Use measures help leaders, decision-makers, and physicians identify improvement opportunities for affordability and value in our healthcare systems. The measures provide transparent information needed to drive change for better health and experience at a lower cost for our patients and communities.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by Independent Health

Independent Health supports endorsement of HealthPartner's Total Cost of Care (#1604 and #1598) measures. These measures have been adopted by many stakeholders in the health care community and have advanced the national discussion on health care affordability.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by Andrew Dorwart, MD

Stillwater Medical Group and Lakeview Hospital is an integrated, non-profit clinic and hospital system serving the eastern Twin Cities metro area and western Wisconsin. We use HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures in our system to identify opportunities to improve affordability for our patients. We support maintaining endorsement of the HealthPartners measures.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by Park Nicollet Health Services

Park Nicollet appreciates the opportunity to voice our support for HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. HealthPartners has transparently shared the measurement method and measure results with providers in our community for nearly a decade, and we have used these measures to improve health care affordability for our patients, while maintaining top quality performance. Our full statement of support and comment on the usability and usefulness of these measures was submitted as part of HealthPartners Total Cost of Care and Total Resource Use NQF submission.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by Sanford Health

Sanford Health supports endorsement of the HealthPartners Total Cost of Care (#1604) and Total Resource Use (#1598) measures. As an integrated health system in the HealthPartners network, we appreciate the transparency and soundness of the measures, as well as our partnership with HealthPartners as we strive to improve care for our patients. The Sanford Health Plan is also a licensee and user of the measures.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by HealthPartners Medical Group

HealthPartners Medical Group strongly recommends for endorsement both the Total Cost of Care (#1604) and Total Resource Use (#1598) measures. For more than a decade, Total Cost of Care (TCOC) has been the top line measure of affordability for our care group. We drill down from the overall measure of TCOC to price drivers, and Total Resource Use drivers to identify opportunities for improvement. These measures have guided our improvement strategies, allowing us to focus on appropriate use of services and place of service opportunities. This has resulted in improved affordability for our patients. Our full statement of support and usability of these measures was included in the measure submission.

1604 Total Cost of Care Population-Based PMPM Index

Submitted by Network for Regional Healthcare Improvement

On behalf of NRHI, we are in support of NQF endorsing this measure. For over three years we have been actively engaged with regions across the country measuring, reporting and using the total resource use population based PMPM index. Recently we published a benchmark report that utilized this measure and compared across 5 regions which has resulted in meaningful conversations within regions about the cause of variation. Seven regions have produced and distributed attributed practice level reports in their communities at least once, some multiple times over the past few years. During 2015, healthcare cost information on over 5 million patients attributed to 20,000 individual physicians were included in practice level reports and used by practices to identify areas of variation and opportunities for intervention to improve care while decreasing costs. This measure allows practices, regions and states to better understand how they compare and gain understanding of the underlying causes of the variation.

The basic foundation for all of these efforts is the HealthPartners NQF endorsed TCOC measure framework. NRHI has been awarded funding from RWJF for a third phase which began on November 1, 2016. During this two-year grant, we will expand the number of regions producing, sharing and using TCOC for both commercial and Medicare populations, maintain and grow our Getting to Affordability Learning Modules and community - a place to connect with others across the country who are measuring and using TCOC, convene a multi-stakeholder summit on using TCOC to advance the Triple Aim and payment reform, and develop and implement sustainability plans to ensure future ability to produce, share and use TCOC.

We support further endorsement of this measure and would be happy to answer any questions.

2158 Medicare Spending Per Beneficiary (MSPB) – Hospital

Submitted by American Medical Association

The AMA continues to remain concerned with the use of this measure. The measure is currently in use within physician programs, but testing has only been performed at the hospital level, which is a serious concern since it cannot be assumed that this measure will have the same impact in a physician practice as in the hospital. We remain concerned over the variation in discharge costs and how much control a hospital has over them. Some hospitals may have a direct connection with a rehab facility and therefore would have some control over the costs associated with rehab. In other instances, a hospital may have no connection or ownership over a rehab facility and based on the availability of space with the non-connected rehab facility and therefore, no true control over the costs associated with rehab or continued relationship.

We also have the following concerns and ongoing questions:

The specifications state that the time period during which costs are captured are 3 days prior to admission to 30 days after discharge. As specified and defined there are instance where more than 30-days cost may be included. For example, if a patient is admitted to an IRF following discharge and that stay is longer than 30 days, all costs from that stay are included. Therefore, if the patient is in the IRF for 45 days, the developer includes 45 days of charges and not 30. We, question whether the length of an IRF stay are truly within the control of a hospital or physician and to what degree the costs beyond the 30-day window impact a hospital's or physician's score. This was not included or examined in the validity testing and it would be helpful for the developer to test the issue.

Reliability Testing: The developer states that data element reliability was completed based on CMS' audit process, but no data is provided to support whether the audit actually yields reliable results. Therefore, we question whether the developer is justified in stating that reliability was completed without any results. We also question the reliability score of 0.4 with 25 episodes. Acumen's previous submission mentioned that they provided the confidence intervals (CIs) but they are not in this current submission and understanding how wide the intervals are would be incredibly helpful with understanding the reliability of the measure. Therefore, we urge Acumen to release this information for the committee and the public to review. During the last review, Acumen stated that if they increased the minimum number of episodes to 50 the number of hospitals included goes from 99% to 95.9%, but did not provide the reliability score with 50 episodes. The AMA believes it is better to have a higher reliability score than capturing the maximum number of hospitals. The low reliability score of 0.4 leads to too much noise with the measure and inaccurate and faulty conclusions about care.

We continue to remain concerned with the data provided for the test/re-test results and the validity of the measure. The test/retest results showed approximately 30% of hospitals in the lowest spending quintile in one sample were not in the lowest spending quintile in the next sample. In addition, 30% of hospitals in the highest spending quintile in one sample were not in the highest spending quintile in the next sample.

Validity Testing: We request further review of the validity testing results from the 2012 submission. In S.11, Interpretation of Scores, it is stated that the measure should not be used alone since the results alone do not necessarily reflect the quality of care provided. Yet, when they tested the correlation of MSPB to the readmission measure (also used in physician programs) last time, CMS found a very weak association between the two. However, Acumen did not do any further testing on the correlation of cost with quality during this current review and given the omission of information it calls into question the usability and validity of the measure.

2158 Medicare Spending Per Beneficiary (MSPB) – Hospital

Submitted by Federation of American Hospitals

The Federation of American Hospitals ("FAH") requests that the Admissions and Readmissions Standing Committee provide input on what new factor(s) and/or new analyses might be needed on measure #2158, "Medicare Spending per Beneficiary (MSBP) - Hospital" in light of the recent report released by the Office of the Assistant Secretary for Planning and Evaluation (ASPE). Specifically, the ASPE report provided further confirmation that sociodemographic factors are strongly linked to hospital performance on resource use. Plus, NQF and this committee should address the potential unintended consequences of continuing to endorse measures without sufficient adjustment. The FAH encourages the committee to request additional analyses from the developer if needed. FAH believes it is critical that the NQF evaluations of measures such as this one continue to factor in new information and recommendations given the constantly evolving nature and understanding of the role of SES.

Appendix F: Cost and Resource Use Measure Evaluation Criteria: Update Recommendations

The purpose of this document is to review the current criteria used by NQF to evaluate candidate cost and resource use measures for endorsement. Over the past several years, experience has been gained from deploying the criteria in multiple CDP projects, in addition to updates to the quality measure evaluation criteria. Staff has made preliminary recommendations on updates to the criteria for the Standing Committee to consider.

Note: Red paragraphs with strikethrough marked with an asterisk (*) are to be removed. Purple paragraphs marked with a double dagger (‡) are to be added.

Importance to Measure and Report

1a. The measure focus addresses:

~~– a specific national health Goal/Priority identified by DHHS or the National Priorities Partnership convened by NQF:*~~

~~OR*~~

– a demonstrated high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality).

AND

Performance Gap

1b. Demonstration of resource use or cost problems and opportunity for improvement, i.e., data demonstrating

- Considerable variation cost or resource across providers; and/or
- Disparities in care across population groups

~~1c. The intent of the resource use measure4 and the measure construct are clearly described. AND*~~

~~The resource use service categories (i.e., types of resources/costs) that are included in the resource use measure are consistent with and representative of the intent of the measure.*~~

2. Scientific acceptability of the measure properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the cost or resources used to deliver care. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

2a. Reliability

2a1. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability.

- All measures that use the ICD classification system must use ICD-10-CM.‡

• eMeasures should be specified in the Health Quality Measures Format (HQMF) and must use the Quality Data Model (QDM) and value sets vetted and published through the National Library of Medicine's Value Set Authority Center (VASC).‡

2a2. Reliability testing demonstrates that the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or that the measure score is precise.

2b. Validity
2b1. The measure specifications are consistent with the measure intent and captures the most inclusive target population.

2b2. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided.

2b3. Exclusions are supported by the clinical evidence
AND/OR

There is a rationale or analysis demonstrating that the measure results are sufficiently distorted due to the magnitude and/or frequency of then on-clinical exclusions;
AND

− Measure specifications for scoring include computing exclusions so that the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
AND

− If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent10 (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b4. For resource use measures and other measures when indicated:
− an evidence-based risk-adjustment strategy is specified and is based on patient factors (including clinical and sociodemographic risk factors) that influence the measured outcome and are present at start of care, and has demonstrated adequate discrimination and calibration
OR
− rationale/data support no risk-adjustment/-stratification.

2b5. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/ clinically meaningful differences in performance.

2b6. If multiple data sources/methods are specified, there is demonstration that they produce comparable results.

2c. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender)
OR
rationale/data justifies why stratification is not necessary or not feasible.

3. Feasibility
Extent to which the required data are readily available or could be captured without undue burden, and can be implemented for performance measurement.
3a. For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3b. The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3c. Demonstration that the data collection strategy (e.g., data source/availability, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) or elements such as risk model, grouper, instrument) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

4. Usability and use
Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency
Performance results are used in at least one accountability application one within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided. AND

4b. Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high quality, efficient healthcare for individuals or populations. AND

4d. Data and result detail are maintained such that the resource use measure, including the clinical and construction logic for a defined unit of measurement can be deconstructed to facilitate transparency and understanding.

National Quality Forum
1030 15th St NW, Suite 800
Washington, DC 20005
http://www.qualityforum.org

ISBN 978-1-68248-070-0