NATIONAL QUALITY FORUM

Memo

April 19, 2017

TO: NQF Members and Public

FR: NQF Staff

RE: Commenting Draft Report: *National Voluntary Consensus Standards for Cost and Resource Use 2016-2017*

Background

This report reflects the review of measures in the Cost and Resource Use project. Healthcare spending in the United States has increased exponentially, yet the United States ranks below other developed countries for quality of care and health outcomes. The goal of this project is to improve efficiency through measurement in order to reduce the rate of cost growth and improve the quality of care provided.

The 19-person Cost and Resource Use Standing Committee reviewed three measures; all three were recommended for endorsement.

Recommended:

- 1598: Total Resource Use Population-based PMPM Index (HealthPartners)
- 1604: Total Cost of Care Population-based PMPM Index (HealthPartners)
- 2158: Medicare Spending per Beneficiary – Hospital (Acumen)

The Committee requests comments on all measures.

NQF Member and Public Commenting

NQF Members and the public are encouraged to provide comments via the online commenting tool on the draft report as a whole, or on the specific measures evaluated by the Cost and Resource Use Standing Committee.

Please note that commenting concludes on May 19, 2017 at 6:00 pm ET – no exceptions.

CDP: Cost and Resource Use Project 2016-2017

DRAFT REPORT FOR COMMENT

April 20, 2017

NATIONAL QUALITY FORUM

This report is funded by the Department of Health and Human Services under contract HHSM-500-2012-00009I Task Order HHSM-500-T0000

Contents

Cost and Resource Use Project 2016-2017

DRAFT REPORT

Executive Summary

In 2015, healthcare spending in the United States (U.S.) reached $3.2 trillion or approximately $9,990 per person.[1] This represented a 5.8% 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 contributing factors to these concerning trends are as complex as the healthcare system itself and include physician practice patterns, regional market influences, and access to care. By improving efficiency, there is potential to simultaneously reduce the rate of cost growth and improve the quality of care provided.

As ongoing health reform efforts focus on expanding coverage, increasing access to care, and reducing costs, it is important to understand the current use of resources in the healthcare system in the context of quality, preferably related 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 value.

For this project, the Standing Committee evaluated three measures undergoing maintenance review against NQF's evaluation criteria. All three measures were recommended for continued endorsement. The three measures that were recommended for continued endorsement by the Standing Committee are:

- 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 currently under review are included in the body of the report; detailed summaries of the Committee's discussion and ratings of the criteria for each measure are in Appendix A.

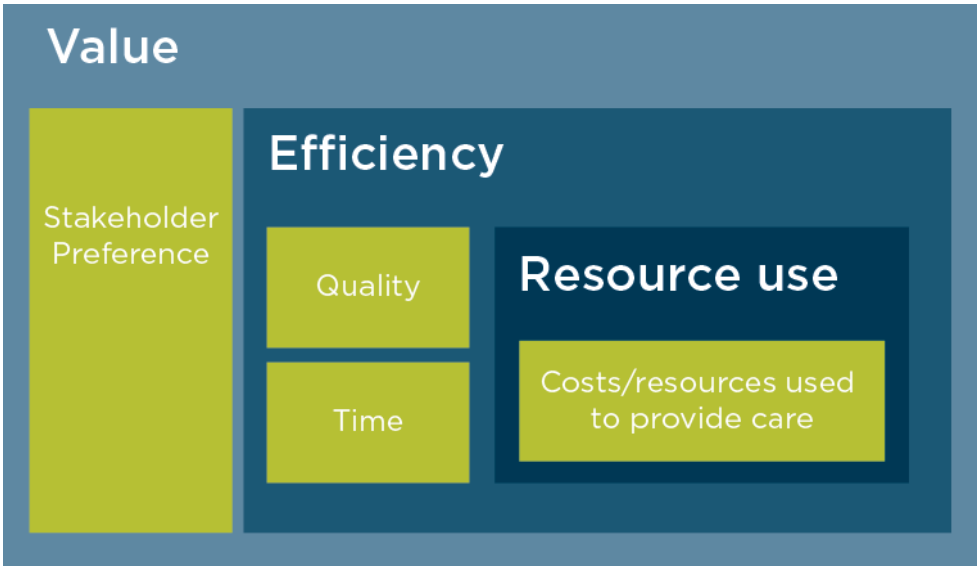
Introduction

In 2015, healthcare spending in the United States (U.S.) reached $3.2 trillion or approximately $9,990 per person.[5] This represented a 5.8% 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 contributing factors to these concerning trends are as complex as the healthcare system itself and include physician practice patterns, regional market influences, and access to care. By improving efficiency, there is potential to simultaneously reduce the rate of cost growth and improve the quality of care provided.

As ongoing health reform efforts focus on expanding coverage, increasing access to care, and reducing costs, it is important to understand the current use of resources in the healthcare system in the context of quality, preferably 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) require 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 utilization. Resource use measures can be defined as measures of health services that are applied to a population or event.[9] A resource use measure counts the frequency 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 price(s), by payor or consumer, for a healthcare service or group of services associated with a specified patient population, time period, and unit(s) 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 non-condition specific measures of total cost, using both per-capita or per-hospitalization episode approaches. This fourth phase involved the review of three non-condition 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 D) oversees NQF's portfolio of six Cost and Resource Use measures (see Appendix B).

Table 1. NQF Cost and Resource Use Portfolio of Measures

NQF #	Title	Category
1598	Total Resource use Population-based PMPM Index	Non-condition specific per capita resource use measure
1604	Total Cost of Care Population-based PMPM Index	Non-condition specific per capita cost measure

NQF #	Title	Category
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	Non-condition 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 health care delivery models.

Use of Measures in the Portfolio

NQF endorsement is valued because the evaluation is rigorous and transparent, and conducted by multi-stakeholder 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 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 their discussions, the Committee identified areas for cost and resource use measure development, including:

- Total per capita cost measure for Medicare patients

- Cost and resource use 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 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 is included in Appendix F. A summary of the proposed changes include:

- 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 sub criterion 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.

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 sub-criteria 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 recommended for endorsement	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%). However, some Committee members urged developers to provide more information on those providers whose measures 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 these outliers 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 a less vulnerable

patient. The Committee stressed the importance of ensuring cost and resource use measures are appropriately risk adjusted to ensure 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, #2158, assesses the total spending per Medicare beneficiary for 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 to not 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 a difficult endeavor. 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 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 and ratings of the criteria for each measure are in included in Appendix A.

1598: Total Resource Use Population-based PMPM Index (HealthPartners): Recommended

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 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 the measure is currently used widely and agreed it was 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 measure #1598 meets the NQF criteria, and recommended it for continued endorsement.

1604: Total Cost of Care Population-based PMPM Index (HealthPartners): Recommended

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**: «measurement_level»; **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 this measure addresses an important aspect of healthcare.

The Committee agreed the measure continues to demonstrate 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. 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 #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 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 measure #1604 met the criteria, and recommended it for continued endorsement.

2158: Medicare Spending Per Beneficiary (MSPB) - Hospital (Acumen, LLC): Recommended

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 was generally supportive of 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 a precise enough measure 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 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 the measure is feasible and widely used, but suggested 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. Ultimately, the Committee agreed the measure met the NQF criteria, and recommended #2158 for continued endorsement.

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*. 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. 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. 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. 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 report: Evaluating Efficiency Across Patient-Focused Episodes of Care
[11] https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/highlights.pdf
[12] https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/highlights.pdf

Appendix A: Details of Measure Evaluation

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

Measures Recommended

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

STEERINGSTANDING COMMITTEE MEETING [03/15/2017]

1598 Total Resource Use Population-based PMPM Index

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(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**

<u>Rationale</u>:

- 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.

1598 Total Resource Use Population-based PMPM Index

- 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:

1598 Total Resource Use Population-based PMPM Index

- 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, there is a multistakeholder committee that provided input on the measure.

1598 Total Resource Use Population-based PMPM Index

- 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 there are multiple accountability programs that this measure is utilized in 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.

1598 Total Resource Use Population-based PMPM Index
6. Public and Member Comment •
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

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

1604 Total Cost of Care Population-based PMPM Index

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: <u>The measure meets the Importance criteria</u>

(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**

<u>Rationale</u>:

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

<u>Rationale</u>:

- 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.

1604 Total Cost of Care Population-based PMPM Index

- The developer reported one change to the measure specifications. Previously, members were 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

1604 Total Cost of Care Population-based PMPM 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).

1604 Total Cost of Care Population-based PMPM Index

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 there are multiple accountability programs that this measure is utilized in 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 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

1604 Total Cost of Care Population-based PMPM Index
•
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

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: <u>The measure meets the Importance criteria</u>** (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;** <u>Rationale</u>: • 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 31st, 2017.

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 highlighted 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 considers 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

-

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

8. Board of Directors Vote: Y-X; N-X

2158 Medicare Spending Per Beneficiary (MSPB) - Hospital
9. Appeals

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: Project 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
Senior Vice President, Quality Measurement

Elisa Munthali, MPH

Vice President, Quality Measurement

Erin O' Rourke
Senior Director

Rachel Roiland, PhD, RN
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.

	1598: Total Resource Use Population-based PMPM Index
	○ 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	Copyright statement: © 2016 HealthPartners. Reprints allowed for noncommercial purposes only if this copyright notice is prominently included and HealthPartners is given clear attribution as the copyright owner. Disclaimer: Total Cost of Care and Total Resource Use are licensed free of charge with supporting implementation tools at the following website: www.healthpartners.com/tcoc

	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)

	1604: Total Cost of Care Population-based PMPM Index
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

	1604: Total Cost of Care Population-based PMPM Index
Copyright / Disclaimer	Copyright statement: © 2016 HealthPartners. Reprints allowed for noncommercial purposes only if this copyright notice is prominently included and HealthPartners is given clear attribution as the copyright owner. Disclaimers: Total Cost of Care and Total Resource Use are licensed free of charge with supporting implementation tools at the following website: www.healthpartners.com/tcoc

	2158: Medicare Spending per Beneficiary (MSPB) – Hospital
Steward	Acumen
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.

	2158: Medicare Spending per Beneficiary (MSPB) – Hospital
	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.
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).

	2158: Medicare Spending per Beneficiary (MSPB) – Hospital
Copyright / Disclaimer	N/A

Appendix E: Pre-Evaluation Comments

Comments received as of March 06, 2017.

Topic	Commenter	Comment
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

Topic	Commenter	Comment
		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.

Topic	Commenter	Comment
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. Thomas Foels, MD, MMM EVP / Chief Medical Officer Independent Health
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. Angelo Sinopoli, MD VP, Clinical Integration, CMO Greenville Health System
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. Paul Kasuba, MD SVP/CMO
1598: Total Resource Use Population-based PMPM Index	Submitted by Sia Lo	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. Nance McClure

Topic	Commenter	Comment
		Chief Operating Officer Brian Rank MD Executive Medical Director Beth Averbeck MD Senior Medical Director Primary Care
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. Andrew Dorwart, MD Stillwater Medical Group President, Lakeview Hospital System CMO
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. Steve Connelly, MD President, Park Nicollet Health Services Kristi Lyon Vice President, Payer Relations
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

Topic	Commenter	Comment
		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

Topic	Commenter	Comment
		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

Topic	Commenter	Comment
		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

Topic	Commenter	Comment
		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. Thomas Foels, MD, MMM EVP / Chief Medical Officer Independent Health
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. Andrew Dorwart, MD Stillwater Medical Group President, Lakeview Hospital System CMO
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. Steve Connelly, MD President, Park Nicollet Health Services Kristi Lyon Vice President, Payer Relations

Topic	Commenter	Comment
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 Sia Lo	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

Topic	Commenter	Comment
		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

Topic	Commenter	Comment
		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)

Topic	Commenter	Comment
		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.

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 [Assigned by Communications]