NATIONAL QUALITY FORUM

Memo

TO: Consensus Standards Approval Committee (CSAC)

FR: Cost and Resource Use Team

RE: Cost and Resource Use Project, 2016-2017

DA: July 11, 2017

CSAC ACTION REQUIRED: The CSAC will review recommendations from the Cost and Resource Use project at its July 11-12, 2017 meeting and vote whether to uphold the recommendations from the Committee.

This memo includes a summary of the project, recommended measures, and themes identified from and responses to the public and member comments.

NQF Member voting on these recommended measures will close on July 5, 2017.

Accompanying this memo are the following documents:

1. **Cost and Resource Use Project 2016-2017 Draft Report.** The draft report has been updated to reflect the changes made following Standing Committee discussion of public and member comments. The complete draft report and supplemental materials are available on the project page.
2. **Comment Table.** Staff has identified themes within the comments received. This table lists 21 comments received during the post meeting comment period and the measure developer, NQF, and Standing Committee responses.

BACKGROUND

The NQF portfolio of Cost and Resource Use Measures specifically addresses the priority of making care more affordable for individuals, families, employers, and the U.S. government. These measures also support the development and spread of new health care delivery models.

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.

DRAFT REPORT

The Cost and Resource Use Draft Report presents the results of the evaluation of three measures considered under the Consensus Development Process (CDP). All three measures are recommended for endorsement.

The measures were evaluated against the current version of the measure evaluation criteria.

	Maintenance	New	Total
Measures under consideration	3	0	3
Measures recommended for endorsement	3	0	3

CSAC ACTION REQUIRED

Pursuant to the CDP, the CSAC is asked to consider endorsement of three candidate consensus measures.

Cost and Resource Use Measures Recommended for Endorsement:

- 1598: Total Resource Use Population-based PMPM Index
  - Overall Suitability for Endorsement: Y-18; N-0
- 1604: Total Cost of Care Population-based PMPM Index
  - Overall Suitability for Endorsement: Y-18; N-0
- 2158: Medicare Spending Per Beneficiary (MSPB) – Hospital
  - Overall Suitability for Endorsement: Y-17; N-1

COMMENTS AND THEIR DISPOSITION

NQF received 21 comments from 9 organizations (including 9 member organizations) and individuals pertaining to the general draft report and to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Cost and Resource Use project page under the Public and Member Comment section.

Comment Themes and Committee Responses

Comments about specific measure specifications and rationale were forwarded to the developers, who were invited to respond.

The Standing Committee reviewed all of the submitted comments (general and measure specific) and developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

Theme 1 – Concerns about Reliability and Validity

A number of commenters questioned the reliability and validity of the measures. Comments on #2158*: Medicare Spending Per Beneficiary (MSPB) – Hospital* raised concerns about the weak association with measures of readmissions as well as the attribution model of the measure. Commenters noted that post-acute spending drives most of the variation in the measure and that hospitals may have limited ability to influence their results. Comments on #1598: *Total Resource Use Population-based PMPM Index* and #1604: *Total Cost of Care Population-based PMPM Index* raised concerns that testing occurred only in two states. Additionally commenters

requested additional details on standardized prices, risk adjustment approaches, and acceptable sample sizes to ensure the measures enable accurate comparisons.

Developer Response:

Response 1: HealthPartners

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Minnesota Community Measurement (MNCM) is a community collaborative organization that measures Total Cost of Care and Total Resource Use annually, according to the NQF-endorsed specifications, for all provider groups in the state of Minnesota and bordering communities to drive improvement by showcasing variation through transparency. The Network for Regional Healthcare Improvement (NRHI) represents more than 30 Regional Health Improvement Collaboratives (RHICs) across

the United States and three state- affiliated partners. With funding from the Robert Wood Johnson Foundation, five pioneering regional health improvement collaboratives (RHICs) are now joined by six additional regions to standardize how they report cost information. NRHI is driving a national effort to make care affordable by using the NQF-endorsed specifications to make cost and resource information consistent and transparent. Both MNCM (third year) and NRHI (first year) results were included in the usability portion of our submission documents for the committee's review.

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

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

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

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

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

data source with more granular income data.

Citations:
1. Alter D, et. al. Lesson From Canada's Universal Care: Socially Disadvantaged Patients Use More Health Services, Still Have Poorer Health.. Health Affairs doi: 10.1377/hlthaff.2009.0669Health Aff February 2011 vol. 30 no. 2 274-283 http://content.healthaffairs.org/content/30/2/274.abstract?sid=94d288f0- 331d-469e-8c11-023b272bed92
2. Lemstra, M, et.al. High health care utilization and costs associated with lower socio-economic status: results from a linked data set. Can J Public Health. 2009 May-Jun;100(3):180-3. https://www.jstor.org/stable/41995241?seq=1#page_scan_tab_contents 4.
3. United States Department of Labor, Bureau of Labor Statistics 2015. Earnings and Unemployment by Educational Attainment Status. Last Modified March 15, 2016. http://www.bls.gov/emp/ep_chart_001.htm

Response 2: Acumen/Centers for Medicare & Medicaid Services

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

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

Variance in provider scores based on post-discharge spending emphasizes the importance of care transitions and care coordination in improving patient care. Hospitals receive a Hospital-Specific Report (HSR) that provides information on the hospital's performance on the MSPB-Hospital measure, as well as three supplementary hospital-specific data files (an index admission file, a beneficiary risk score file, and an MSPB-Hospital episode file) related to the hospital's MSPB-Hospital measure. Together,

these files provide an overview of how the hospital performed on the MSPB-Hospital measure and identify other providers involved in care for their beneficiaries, which facilitates better coordination of care with those providers. No evidence of unintended consequences to individuals or populations, such as changes in referral patterns, have been identified during testing and since implementation.

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

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

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

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

Response 3: Acumen/Centers for Medicare & Medicaid Services

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

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

Variance in provider scores based on post-discharge spending emphasizes the importance of care transitions and care coordination in improving patient care.

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

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

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

suggests that these differences may be driven by hospitals with a very high concentration of dual eligible beneficiaries (above 60%), and that measure scores are high for both duals and non-duals in these hospitals. This suggests that these hospitals are relatively higher-cost hospitals for all types of patients.

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

Committee Response:
The Committee has reviewed your comment and appreciates your input. The Committee recognizes the need to ensure NQF-endorsed cost and resource use measures are reliable and valid.

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

The Committee noted that #1598 and #1604 have been widely implemented and users have supported the usefulness of the information generated by the measures.

Theme 2 – Adjusting for Social Risk Factors

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

Developer Response:
Response 1: HealthPartners
HealthPartners thanks the American Medical Association (AMA) for sharing its comments.

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

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

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

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

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

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

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

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

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

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

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

The HealthPartners measures of Total Cost of Care and Total Resource Use are not

measures of value, they are measures that represent the affordability arm of the Triple Aim. HealthPartners is focused on the Triple Aim, quality, experience and affordability of health care to provide value for our patients. We believe it is essential when measuring affordability to complement it with quality and experience measurement, which allows members and patients to make their own value determination. The majority of the contracted provider groups in our health plan network are focusing on the same work, each having their own process improvement plan. HealthPartners' health plan implemented a Triple Aim risk sharing program over 5 years ago with Total Cost of Care representing the affordability component of the Triple Aim. To date, tens of millions of dollars have been paid out by the health plan and self-insured employer groups. Essentia Health, CentraCare Health and Fairview Health Services are provider groups outside the HealthPartners family of providers who have shared their success stories through letters of support in the usability section of the National Quality Forum measure submission process.

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

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

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

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

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

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

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

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

Citations:

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

2. Lemstra, M, et.al. High health care utilization and costs associated with lower socio-economic status: results from a linked data set. Can J Public Health. 2009 May-Jun;100(3):180-3.

https://www.jstor.org/stable/41995241?seq=1#page_scan_tab_contents 4.
3. United States Department of Labor, Bureau of Labor Statistics 2015. Earnings and Unemployment by Educational Attainment Status. Last Modified March 15, 2016. http://www.bls.gov/emp/ep_chart_001.htm

Response 2: Acumen/Centers for Medicare & Medicaid Services

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

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

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

The developer would also like to note that they submitted an updated measure testing form to the NQF on March 31st, 2017 that contained an appendix with additional analyses responding to NQF feedback and further description of the original submission. That appendix notes that analyses comparing the MSPB-Hospital measure with the condition-specific readmission measures were excluded in the 2016 submission because

the condition-specific readmission measures examine hospital performance on a specific set of conditions, while the MSPB-Hospital measure is intended to capture hospital performance across all acute conditions. Consequently, comparisons could be misleading. Since MSPB-Hospital is an all cost measure that includes all conditions, the developer believes that it is more appropriate to look at the correlation between MSPB-Hospital and another broad-based all cost measure (i.e., the risk-adjusted, standardized total Medicare spending at the Hospital Referral Regions (HRR) level). The developer agrees that the MSPB-Hospital measure is most meaningful when presented in the context of other quality measures, which are part of the Hospital Value-Based Purchasing (VBP) Program. As part of the Hospital VBP Program, the MSPB-Hospital measure is combined with current quality of care measures to facilitate profiling hospital value (payments and quality).

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

The MSPB-Hospital measure developer appreciates the commenter's feedback on the separate clinician-level measure (MSPB-TIN) used in the Merit-Based Incentive Payment System (MIPS). The developer would like to clarify that while MSPB-TIN and the facility-level MSPB-Hospital measure currently under consideration for NQF re-endorsement are alike, MSPB-TIN differs in attribution methodology. The MSPB-TIN measure is still under reevaluation. To ensure the reliability and validity of the measures being implemented, CMS reevaluates the measures annually and plans NQF submission of the measures by taking into account program needs and measure implementation timelines to meet the statutory requirements. MSPB-TIN was finalized for inclusion in the MIPS Cost Category as part of the Quality Payment Program Final Rule. The first performance

period for the Cost Category is calendar year 2017, and the category is weighted at zero percent for the associated payment year, meaning that it will not impact payments under the program in its first year.

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

Response 3: Acumen/Centers for Medicare & Medicaid Services

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

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

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

The developer would also like to note that they submitted an updated measure testing

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

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

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

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

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

Theme 3 – Concerns About Populations Included in the Measures

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

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

> *Developer Response:*
> **Response 1: HealthPartners**
> HealthPartners' agrees that cost measures are not a marker of quality, and should not be used to draw conclusions on quality.
>
> As a provider of cancer care, (American Society of Clinical Oncology-certified practices HealthPartners Regions Hospital Cancer Care Center and Park Nicollet Frauenshuh Cancer Center), HealthPartners' understands the complexity of cancer diagnosis, the importance of early detection and the variation in treatments as we care for cancer patients in our medical groups.
>
> We are sensitive to the complexity of including cancer patients in our full population

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

Response 2: Acumen/Centers for Medicare & Medicaid Services

Thank you for your comment. We recognize that cancer patients often have complex comorbidities and require more intensive treatment. The MSPB-Hospital measure accounts for comorbidities through risk adjustment. Specifically, the MSPB-Hospital risk adjustment methodology adjusts the MSPB-Hospital measure for age, severity of illness, and enrollment status indicators. The methodology includes 12 age-categorical variables; 79 hierarchical condition category (HCC) variables derived from the beneficiary's claims during the period 90 days prior to the start of the episode to measure severity of illness; as well as the MS-DRG of the index hospitalization. The risk adjustment methodology also includes the HCC interaction variables, status indicator variables for whether the beneficiary qualifies for Medicare through Disability or End-Stage Renal Disease (ESRD), and whether a beneficiary resides in a long-term care facility. The HCC variables include several HCCs that indicate the presence of cancer, as well as an interaction term for cancer and immune disorders. These HCCs are included to capture the variation in cost of care that beneficiaries with cancer may have. The MS-DRG is included in the risk adjustment model to better account for the differences in cost of care that stem from different reasons for hospitalization, including cancer. This allows the MSPB-Hospital measure to compare cost of care across all conditions, rather than focusing on a specific disease. As such, the risk factor of the MS-DRG of the index hospitalization should account for the more intensive treatment that cancer patients may require. Additionally, to mitigate the effect of high-cost and low-cost outliers on each hospital's MSPB-Hospital measure score, outliers are excluded at the episode level. Specifically, MSPB-Hospital episodes whose residuals fall above the 99th percentile or below the 1st percentile of the distribution of residuals across all MSPB-Hospital episodes are excluded from the MSPB-Hospital calculation. Excluding outliers based on residuals eliminates the episodes that deviate most from their predicted values in absolute terms.

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

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

Response 3: HealthPartners

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

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

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

The measures are population-based, patient-centered and cross all categories of health care services, including pharmacy. Pharmacy contributes an estimated 20% of the total costs and resources and are driving steep trends. Therefore it is imperative to include pharmacy costs when measuring total cost of care and resource use. At HealthPartners

we have and are enhancing our approaches to make the costs of drugs available to providers and consumers so the decision-making process can be fully informed. We would encourage others to do the same. While a provider may not have control over formulary drug lists, providers do have an opportunity to help educate patients on alternative drug options when there is clinical equivalence yet a large cost difference. Alternative therapies, generics or less expensive brand drugs may be options for patients that providers can help coordinate, leading to lower overall costs, lower out of pocket costs for patients, increased patient experience and most importantly, a better chance consumers can afford to fill and take their medicine as prescribed. Providers also have the ability to manage potential overuse of medications which not only reduces costs, but also improves quality of care for patients.

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

Theme 4 – Support for Measures

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

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

In addition, one general comment was received, noting the gap in measures in this area and supporting the continued endorsement of these three measures.

Committee Response:
Thank you for your comment

Theme 5 – Updates to the Cost and Resource Use Measure Evaluation Criteria

Overall, commenters were supportive of the revisions to the Cost and Resource Use Measure Evaluation Criteria. However, commenters asked for additional clarifications on what information should be provided by developers to address the performance gap subcriterion.

NQF Response:
NQF thanks the commenters for their support for the revisions to the Cost and Resource Use Measure Evaluation Criteria. The performance gap subcriterion is meant to address the question of whether there is actually a cost and resource use problem that is addressed by a particular measure. Because the measurement enterprise is resource intensive, NQF's position is to endorse measures that address areas of known gaps in performance (i.e., those for which

there is actually opportunity for improvement). Opportunity for improvement can be demonstrated by data that indicate overall poor performance (in the activity or outcome targeted by the measure), substantial variation in performance across providers, or variation in performance for certain subpopulations (i.e., disparities in care). The proposed update removes subcriterion 1c to streamline the criteria, harmonize with the quality measure evaluation criteria, and prevent redundancies with the reliability and validity subcriterion.

NQF MEMBER VOTING RESULTS

NQF Member Vote will close on July 5. Voting results will be sent to the CSAC in an addendum on July 7.

Appendix A – Measure Evaluation Summary Tables

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

1598 Total Resource Use Population-based PMPM Index

Data Source: Claims (Only)

Measure Steward: HealthPartners

STEERINGSTANDING 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; 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**

1598 Total Resource Use Population-based PMPM Index

Rationale:

- This per capita (population- or patient-based) measure calculates total resource use associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services and is expressed as a ratio.
- To interpret, a score greater than 1.00 indicates higher risk adjusted resource use, compared to a peer group average; a score less than 1.00 indicates less risk adjusted resource use, compared to a peer group average.
- The developer defines peer groups as a group of members, providers, geographic regions or any grouping of member data. The resource use measure will return a value that will be relative to the peer group average (e.g., 1.10 = 10% higher than the peer group average).
- The numerator is calculated as the sum of (Total Medical TCRRV / Medical Member Months) + (Total Pharmacy TCRRV / Pharmacy Member Months). The Johns Hopkins Adjusted Clinical Grouper (ACG) risk score is the measure's denominator.
- To demonstrate **measure score reliability**, the developer compared actual measure scores to scores calculated by two sampling methods:
    - Bootstrapping
    - A 90% random sample
    - The variances from Actual RUI ranged from -0.0036 to 0.0065 in the bootstrap to -0.0020 to 0.0015 in the 90% sample.
- Some Committee members expressed concerns with the reliability testing, noting concerns about the attribution approach used in the testing of the measure. The developer responded they used the attribution method used within their health plans.
- The Committee was also concerned that testing only occurred in a localized area (with data from one payer and limited geographic area), raising concerns that the results may not be generalizable and not applicable on a nationwide scale. However, this measure is currently widely used across the country and has demonstrated reliability among other users.
- Ultimately, the Committee agreed the measure met the reliability criterion.
- For this maintenance submission, the developer summarized updated validity testing conducted using provider data from 2014 and 2015. The validity and reliability testing of the measures was conducted with HealthPartners' commercial population of 470,000 members. This updated validity testing consisted of correlations of the measure components (i.e., ACG scores, unadjusted costs) and measure score with other markers of utilization.
- The developers tested the validity of the underlying data elements and performed empirical validity testing of the measure score.
- To demonstrate **data element validity**, the developer conducted a series of correlation analyses:
    - Measure components (i.e., ACG scores & Non-Risk Adjusted Total Cost Relative Resource Values (TCRRVs))
        - ACG Risk-adjusted Total Cost Index (i.e., the measure score)
        - ACG risk-adjusted Resource Use Index (RUI) (i.e., measure 1598)
        - Non-risk adjusted Total Cost Relative Resource Values (TCRRVs)
        - Price
    - Measure component - Non-Risk Adjusted TCRRVs with non-risk adjusted rates of utilization:
        - Inpatient Admits per 1,000
        - ER per 1,000
        - Outpatient surgery per 1,000
        - High Tech Radiology per 1,000
        - E&Ms per 1,000

        - Lab/Path per 1,000
        - Standard radiology per 1,000
        - Pharmacy per 1,000
        - Measure Components with Composite Utilization
- The developer noted that there is high correlation of the measure components to one another and each component's correlation with the non-risk adjusted TCRRVs as sufficient evidence for the validity of the measure components.
    - The correlation between the non-risk adjusted PMPM and the ACG risk adjusted RUI is 0.45.
- The developer attributes the low correlation between ACG and Price to fact that ACG is an estimate of expected resource use whereas price is the unit cost of services actually provided
- To demonstrate **measure score validity**, the developer conducted a series of correlation analyses:
    - ACG Risk-adjusted Risk Use Index (i.e., the measure score) with:
        - Hospital based Total Cost of Care Index
        - Professional Total Cost of Care Index
        - Pharmacy Total Cost of Care Index
        - ACG risk-adjusted Total Cost Index (i.e., measure 1604)
        - Total Price
    - Service Category RUI (i.e., Inpatient, Outpatient, Professional, Pharmacy) with risk-adjusted service category metrics:
        - Inpatient admit rate
        - ER count
        - Outpatient surgery
        - High tech Radiation
        - E&M Visits
        - Lab/Path
        - Standard Radiology
        - Prescription (Rx) Count
    - Measure Score with Composite Utilization
    - Measure Score Over Time
- The risk adjustment approach utilized in the measure is the Johns Hopkins Adjusted Clinical Grouper (ACG) method, which adjusts for age, gender, and diagnosis (i.e., clinical risk). A conceptual rationale for this risk adjustment approach is provided.
- The risk adjustment approach involves:
    - Grouping International Classification Diagnosis (ICD) diagnosis codes into 32 diagnosis groups (i.e., Aggregated Diagnosis Groups (ADGs)). These ADGs are clinically similar and expected to have similar need for healthcare resources.
    - Adjusted Clinical Groups (ACGs) are created from the ADG assignments and are defined by morbidity, age, and sex. Individual members are then assigned to a single ACG category, which quantifies their risk.
- Individual member ACG weights: Individuals are assigned to an ACG actuarial cell that has a corresponding weight reflecting relative illness burden. The ACG weight is then multiple by their number of eligible member months.
- Providers' ACG Scores are calculated as the sum of their attributed members ACG weights.
- To examine the impact of SDS on the measure scores, the developers used two measures of income – 1) tract-level income, obtained from U.S. Census Tract data, and 2) household-level, obtained from a commercially licensed consumer database purchased by HealthPartners.

1598 Total Resource Use Population-based PMPM Index

- 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.
- 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:

1598 Total Resource Use Population-based PMPM Index

- 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.

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

Comments Received

- This measure received a total of five comments. Two supported continued endorsement, and one comment was about the inclusion of non-generalist OB/GYNs and pharmaceutical resources. The remaining two comments focus on concerns around the measure's testing and usability in states outside of those two the measure was tested in. The commenters requested more detail on the measure's current performance and implementation experiences, given that it has been in use for five years. Specific concerns raised include unintended consequences, standardized prices, a risk adjustment approach, and acceptable sample sizes. There are also additional comments focusing on concerns with SDS. One comment was received on the general draft report requesting clarification on how all three measures address cancer patients. The commenter noted that there can be significant variation in treatment needs, comorbidities, and patient preferences that can influence cost and resource use.

Developer Responses

Response 1: Testing and SDS Concerns

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

To address the AMA's first comment regarding standardized pricing, the Total Resource Use measure uses the Total Care Relative Resource Values (TCRRVs). TCRRVs are a grand linear

scale of relative values designed to evaluate resource use across all types of medical services, procedures and places of service. TCRRVs are based on industry standard weighting systems (RVU, DRG, APC). The values are independent of price and can be used to evaluate providers, hospitals, physicians and health plans against their peers on their efficiency of resource use in treating like conditions.

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

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

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

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

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

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

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

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

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

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

their success stories through letters of support in the usability section of the National Quality Forum measure submission process.

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

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

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

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

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

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

To ensure the study population included lower income ranges, HealthPartners Medicaid population was included along with the full commercial book of business for testing. The Medicaid population has a different reimbursement rate (typically significantly lower) than

the commercial reimbursement rate, which would result in a lower total cost of care. The Medicaid population was included to prove that the product delineation between Medicaid and Commercial sufficiently controls for the variation in cost and resource use, therefore, adding income in the model resulted in no additional explanatory power.

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

Citations:

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

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

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

Response 2: Inclusion of OB/GYNs and Pharmaceutical Resources

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

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

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

1598 Total Resource Use Population-based PMPM Index

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

Response 3: Inclusion of Cancer Patients

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

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

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

Committee Responses

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

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

While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data

1598 Total Resource Use Population-based PMPM Index
elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee recognized the impact that social risk can have on cost and resource use measures and encourages measure developers to test the impact of additional social risk variables. The Committee also encouraged exploration of the impact of community-level variables. However, the Committee generally agreed that the risk adjustment method used in these measures met the NQF criteria given the data available to the developer, and the measure testing results presented. After reviewing the comments received, and the developer's response, the Committee does not wish to reconsider its recommendations on any of the three measures. While the specific attribution model is currently outside of the endorsement criteria, the Committee recognizes the need for further testing of attribution models.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-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

1604 Total Cost of Care Population-based PMPM Index

performance and allow for equitable comparisons of utilization and outcomes. ACGs are defined by morbidity, age, and sex and are person-focused to categorize patients' illnesses. Based on the pattern of morbidities, the ACG approach assigns each individual to a single ACG category.

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

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

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

Type of Measure: Cost/Resource Use

Data Source: Claims (Only)

Measure Steward: HealthPartners

STANDING COMMITTEE MEETING [03/15/2017]

1. Importance to Measure and Report: <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**

1604 Total Cost of Care Population-based PMPM Index

Rationale:

- This per capita (population- or patient-based) measure calculates the total cost of care of a commercial population and is expressed as a ratio.
- To interpret, a score greater than 1.00 indicates a higher paid risk adjusted PMPM value, compared to a peer group average; a score less than 1.00 indicates less paid risk adjusted PMPM value, compared to a peer group average.
- The choice of a peer group is at the discretion of the measure user and can include the internal medicine, family medicine, pediatrics, geriatrics, and OB/GYN specialties and physician, physician assistant, and nurse practitioner provider types. The peer group's average is set at the benchmark.
- The numerator is calculated as the sum of (Total Medical Cost / Medical Member Months) + (Total Pharmacy Cost / Pharmacy Member Months).
- The Johns Hopkins Adjusted Clinical Grouper (ACG) risk scores constitutes the measure's denominator.
- The developer reported one change to the measure specifications. Previously, 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
    - 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.

1604 Total Cost of Care Population-based PMPM Index

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

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

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

Rationale:

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

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

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

Rationale:

1604 Total Cost of Care Population-based PMPM Index

- 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: April 20, 2017 – May 19, 2017

Comments Received

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

Developer Responses

Response 1: Testing and SDS concerns

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

To ensure the study population included lower income ranges, HealthPartners Medicaid population was included along with the full commercial book of business for testing. The Medicaid population has a different reimbursement rate (typically significantly lower) than the commercial reimbursement rate, which would result in a lower total cost of care. The Medicaid population was included to prove that the product delineation between Medicaid

and Commercial sufficiently controls for the variation in cost and resource use, therefore, adding income in the model resulted in no additional explanatory power.

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

Citations:

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

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

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

Response 2: Inclusion of OB/GYNs and Pharmaceutical Resources

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

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

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

1604 Total Cost of Care Population-based PMPM Index

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

Response 3: Inclusion of Cancer Patients

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

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

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

Committee Response

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

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

1604 Total Cost of Care Population-based PMPM Index
While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee recognized the impact that social risk can have on cost and resource use measures and encourages measure developers to test the impact of additional social risk variables. The Committee also encouraged exploration of the impact of community-level variables. However, the Committee generally agreed that the risk adjustment method used in these measures met the NQF criteria given the data available to the developer, and the measure testing results presented. After reviewing the comments received, and the developer's response, the Committee does not wish to reconsider its recommendations on any of the three measures. While the specific attribution model is currently outside of the endorsement criteria, the Committee recognizes the need for further testing of attribution models).
7. Consensus Standards Approval Committee (CSAC) Vote: Y-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

2158 Medicare Spending Per Beneficiary (MSPB) - Hospital

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).

2158 Medicare Spending Per Beneficiary (MSPB) - Hospital

- 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: April 20, 2017 – May 19, 2017

Comments Received

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

Developer Response

Response 1: SDS Adjustment

The measure developer appreciates the AMA's feedback on the MSPB-Hospital measure construction and the testing of sociodemographic (SDS) factors in the measure's risk adjustment model. The developer believes that the MSPB-Hospital measure does meet the

scientific acceptability criteria of validity, and the NQF committee agreed that the measure met the Scientific Acceptability criterion. The NQF committee had 4 members vote that the measure demonstrated high validity, 9 members vote that the measure demonstrated medium validity, and 5 members vote that the measure had low reliability.

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

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

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

The developer believes that the MSPB-Hospital measure submission did meet the requirements of the NQF's SDS trial period and the NQF committee confirmed this by passing the MSPB-Hospital measure on the Scientific Acceptability criterion. The developer noted in the original submission that the inclusion of SDS factors (i.e., family income-to-poverty ratio and race) had a minimal impact on hospital's measure scores. The developer recognizes the commenter's concerns that additional factors could be included in the SDS measure testing. The developer selected family income-to-poverty ratio to strike a balance between the

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

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

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

Response 2: Measure Testing Concerns

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

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

history and MS-DRG. Specifically, the MSPB-Hospital risk adjustment methodology adjusts the MSPB-Hospital measure for age, severity of illness, and enrollment status indicators.

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

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

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

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

Response 3: -Clinician-Level Measurement

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

Response 4: Inclusion of Cancer Patients

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

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

purposes to allow hospitals to evaluate its episode spending before, during, or after the index hospital admission.

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

Committee Response

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

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

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

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

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

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

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

9. Appeals

Cost and Resource Use Project 2016-2017

Consensus Standards Approval Committee Review and Recommendations

July 11-12, 2017

COMMITTEE CO-CHAIRS
Brent Asplin, Co-chair
Cheryl Damberg, Co-chair

NQF STAFF:
Erin O'Rourke, Senior Director
Ashlie Wilbon, Senior Director
Taroon Amin, Consultant
Suzanne Theberge, Senior Project Manager
Hiral Dudhwala, Project Manager
Irvin Singh, Project Analyst

NATIONAL QUALITY FORUM

Cost and Resource Use Project 2016-2017

- This project reviewed three non-condition specific maintenance measures:
  - 2158: Medicare Spending per Beneficiary-Hospital(episode-based cost measure)
  - 1598: Total Resource Use Population-based PMPM Index (per capita resource use measure)
  - 1604: Total Cost of Care Population-based PMPM Index (per capita cost measure)

Cost and Resource Use Project 2016-2017

- Standing Committee recommendations:
  - Recommended all three maintenance measures for re-endorsement

Cost and Resource Use Project 2016-2017

	Maintenance Measures	New Measures	TOTAL Measures
Submitted	3	0	3
Measures Recommended	3	0	3
Measures Not Recommended	0	0	0
Measures Withdrawn from Consideration	0	0	0
Reasons for not recommending:	Importance – X Scientific Acceptability -X Overall – X	Importance – X Scientific Acceptability - X Overall – X	

Overarching Issues

- Risk Adjustment for Social Risk Factors:
  - The Committee discussed the need to test social risk factors capturing individual level attributes and potentially community-level attributes
  - The Committee discussed the impact of adjustment for social risk factors on different groups of providers
  - 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

Overarching Issues (continued)

- Attribution:
  - 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
  - The need to attribute costs must be balanced with the risk of unintended consequences
  - The need for attribution models that support care coordination and team-based care as the system aims to transition from FFS to population-based payment
  - Attribution models should better capture the role of NPs and PAs as a way to address the transition to team-based care

Comments Received

Comments Received:

- 21 post comments from 9 member organizations
- 5 themes were identified:
  - Concerns about reliability and validity
  - Adjusting for social risk factors
  - Concerns about populations included in the measures
  - Support for measures
  - Updates to the Cost and Resource Use Measure Evaluation Criteria

Comments Received: #1598: Total Resource Use Population-based PMPM Index and #1604: Total Cost of Care Population-based PMPM Index

- Concerns around the testing and usability outside of the two testing states, specifically unintended consequences, standardized prices, a risk adjustment approach, and acceptable sample sizes;
- Concerns with the lack of adjustment for social risk factors
- Inclusion of non-generalist OB/GYNs and pharmacy costs
- Support for the measure

Developer Responses: Comments on *#1598: Total Resource Use Population-based PMPM Index* and *#1604: Total Cost of Care Population-based PMPM Index*

- RE: Concerns around the testing and usability in states outside of the two testing states, specifically unintended consequences, standardized prices, a risk adjustment approach, and acceptable sample sizes:
  - The Total Resource Use measure uses the Total Care Relative Resource Values (TCRRVs) which is used to evaluate providers, hospitals, physicians, and health plans against their peers on their efficiency of resource use
  - The measures are specified to use the Johns Hopkins' Adjusted clinical Groups (ACG System); others can be used but must be tested first. To ensure comparability, users must use the same tool.
  - The measure is specified as a full population measure, including all care, from all provider specialties
  - The risk adjustment accounts for variation in age, gender, and clinical risk of patients
  - Income did not specifically impact a patient's total cost or resource use
  - Refer to Comment Table for details of developer response

Developer Responses: Comments on *#1598: Total Resource Use Population-based PMPM Index* and *#1604: Total Cost of Care Population-based PMPM Index*

- RE: Inclusion of non-generalist OB/GYNs and pharmacy costs:
  - The intent of the Total Cost of Care measure is to measure a provider's risk adjusted cost effectiveness at managing the population they care for
  - Implementation of the measures and how results are used, the method of the attribution, the peer group to be used in their respective market, and reporting methods are decisions that need to be considered and defined by the users
  - HealthPartners' measurement approach excludes specific OB/GYN specialties from attribution when measuring primary care providers (e.g. Gynecological Surgery). The measures are however versatile and could evaluate subspecialties if the peer group was limited to the subspecialty being evaluated.
  - Pharmacy contributes an estimated 20% of the total costs and resources and are driving steep trends. Therefore it is imperative to include pharmacy costs when measuring total cost of care and resource use.
  - Refer to Comment Table for details of developer response

Developer Responses: Comments on *#1598: Total Resource Use Population-based PMPM Index* and *#1604: Total Cost of Care Population-based PMPM Index*

- RE: Concerns about inclusion of cancer patients:
  - "We are sensitive to the complexity of including cancer patients in our full population measure and through our testing believe the measures sufficiently adjust for cancer patients in the population through the clinical risk adjustment process and application of the measure criteria."
  - Refer to Comment Table for details of developer response

Comments Received: #2158: Medicare Spending per Beneficiary-Hospital

- Concerns that the measure is only validated and endorsed at the facility level, and not physician level;
- Concerns with the measure's testing for reliability and validity
- Concerns with the lack of adjustment for social risk factors
- Concerns that the majority of variation in the measure is driven by post-acute spending. Commenters noted this measure is used in the Hospital Value-Based Purchasing Program and that there is a potential for negative unintended consequences from its use
- Support for the measure

Developer Responses: Comments on *#2158: Medicare Spending per Beneficiary-Hospital*

- RE: Concerns that the measure is only validated and endorsed at the facility level, and not physician level
  - The developer clarified that the MSPB-Hospital measure (NQF #2158) is not slated for use in the MIPS program, and this NQF submission does not cover the clinician-level MSPB-TIN measure.

Developer Responses: Comments on *#2158: Medicare Spending per Beneficiary-Hospital*

- RE: Concerns with the measure's testing for reliability and validity:
  - The developer responded by addressing rationale for the methodology of the variance and the risk adjustment of the measure.
  - Refer to Comment Table for details of developer response.

Developer Responses: Comments on *#2158: Medicare Spending per Beneficiary-Hospital*

- RE: Concerns with the lack of adjustment for social risk factors:
  - The developer responded by re-clarifying the adjustment for social risk factors done during the original measure submission, as well as the additional analyses requested by the NQF Committee to exam the impact of including a dual eligibility flag in the risk adjustment model.
  - Refer to Comment Table for details of developer response.

Developer Responses: Comments on *#2158: Medicare Spending per Beneficiary-Hospital*

- RE: Concerns that the majority of variation in the measure is driven by post-acute spending and by potential for negative unintended consequences from its use:
  - The developer responded by providing clarification that 84% of the variance in episode cost is accounted for by post-acute care costs, rather than 84% of total episode costs being attributed to the hospital during the 30 day post-discharge period.
  - The developer also emphasized variance in provider scores based on post-discharge spending emphasizes the importance of care transitions and care coordination in improving patient care.
  - Refer to Comment Table for details of developer response.

Developer Responses: Comments on *#2158: Medicare Spending per Beneficiary-Hospital*

- RE: Concerns about inclusion of cancer patients:
  - The developer responded by addressing that the MSPB-Hospital measure accounts for comorbidities through risk adjustment. Specifically, the MSPB-Hospital risk adjustment methodology adjusts the MSPB-Hospital measure for age, severity of illness, and enrollment status indicators.
  - Refer to Comment Table for details of developer response.

Project Timeline and Next Steps

Process Step	Timeline
Appeals Period	July 14-August 14, 2017
Adjudication of Appeals	August 15-September 12, 2017
Final Report	September 26, 2017

Questions?

NATIONAL QUALITY FORUM

TO: Consensus Standards Approval Committee (CSAC)

FR: Cost and Resource Use Team

RE: Cost and Resource Use Project, 2016-2017: Addendum-Member Voting Results

DA: July 7, 2017

The CSAC will review recommendations from the Cost and Resource Use project at its July 11-12, 2017 in-person meeting. This serves as an addendum to the previous memo and contains the updated voting results; the NQF member voting period closed on July 5, 2017.

NQF MEMBER VOTING RESULTS

All of the recommended measures were approved with 86% approval or higher. Representatives of 14 member organizations voted; no votes were received from Health Plan Council. Results for each measure are provided below. (Links are provided to the full measure summary evaluation tables.)

NQF Member Council	Voting Organizations	Eligible to Vote	Rate
Consumer	2	38	5%
Health Plan	0	21	0%
Health Professional	1	104	1%
Provider Organizations	2	110	2%
Public/Community Health Agency	3	15	20%
Purchaser	3	22	14%
QMRI	1	74	1%
Supplier/Industry	2	35	6%
All Councils	**14**	**419**	**6%**

1598: Total Resource Use Population-based PMPM Index]

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1		1	2	100%
Health Plan				0	
Health Professional	1			1	100%
Provider Organizations	2			2	100%
Public/Community Health Agency	3			3	100%
Purchaser	3			3	100%
QMRI	1			1	100%
Supplier/Industry	1	1		2	50%
All Councils	**12**	**1**	**1**	**14**	**92%**
Percentage of councils approving (>60%)					86%
Average council percentage approval					93%

*equation: Yes/ (Total - Abstain)

Voting Comments:

- Minnesota Community Measurement: MN Community Measurement supports the continued endorsement of this total resource use measure. HealthPartners's excellent work in the development of this measure provided a common methodology that can be used for collaboratively reporting total cost and resource use in our community. Our whole hearted support is additionally demonstrated by the implementation of this measure and providing an annual report for public use at http://mncm.org/cost-reports/.
- The Alliance: Our organization uses this measure and we've found the information generated by it to be useful to us and to organizations with which we contract. We've also been very impressed by the transparency and responsiveness of the measure developer whenever we had any questions regarding the measure.

1604: Total Cost of Care Population-based PMPM Index

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	2			2	100%
Health Plan				0	
Health Professional	1			1	100%
Provider Organizations	2			2	100%
Public/Community Health Agency	3			3	100%
Purchaser	3			3	100%
QMRI	1			1	100%
Supplier/Industry	1	1		2	50%
All Councils	**13**	**1**	**0**	**14**	**93%**
Percentage of councils approving (>60%)					86%
Average council percentage approval					93%

*equation: Yes/ (Total - Abstain)

Voting Comments:

- Minnesota Community Measurement: MN Community Measurement supports the continued endorsement of this total cost of care measure. HealthPartners's excellent work in the development of this measure provided a common methodology that can be used for collaboratively reporting total cost of care in our community. Our whole hearted support is additionally demonstrated by the implementation and annual state-wide public reporting of this measure on www.MNHealthscores.org.

- The Alliance: Our organization uses this measure and we've found the information generated by it to be useful to us and to organizations with which we contract. We've also been very impressed by the transparency and responsiveness of the measure developer whenever we had any questions regarding the measure.

2158: Medicare Spending Per Beneficiary (MSPB) - Hospital

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	2			2	100%
Health Plan				0	
Health Professional	1			1	100%
Provider Organizations	1		1	2	100%
Public/Community Health Agency	1		2	3	100%
Purchaser	3			3	100%
QMRI	1			1	100%
Supplier/Industry	1	1		2	50%
All Councils	**10**	**1**	**3**	**14**	**91%**
Percentage of councils approving (>60%)					86%
Average council percentage approval					93%

*equation: Yes/ (Total - Abstain)

Voting Comments:

- No comments received.