NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 1598

De.2. Measure Title: Total Resource Use Population-based PMPM Index

Co.1.1. Measure Steward: HealthPartners

De.3. Brief Description of Measure: 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.

IM.1.1. Developer Rationale: By measuring population based relative resource use, health plans and providers can improve the affordability of health care without sacrificing quality. HealthPartners' RUI gives provider groups valuable information on resource use and, when viewed in conjunction with quality metrics, information on the efficiency of care. The HealthPartners RUI measure is a population-based, patient-centered, total resource use measure, created with Total Care Relative Resource Values that cross all categories of health services. This is in contrast to the many, episodic based resource use measures available in the market today. Both population based and episodic based resource use measures are important and complementary but a key benefit of population based measures is helping to better understand potential overuse & underuse (e.g., although efficient at spine surgery, may be performing too many).

De.1. Measure Type: Cost/Resource Use

S.5. Data Source: Claims

S.3. Level of Analysis: Clinician : Group/Practice, Population : Community, County or City

IF Endorsement Maintenance – Original Endorsement Date: Jan 31, 2012 **Most Recent Endorsement Date:** Jul 13, 2017

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. High impact or high resource use:
The measure focus addresses:
– a demonstrated high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality).
AND
1b. Opportunity for Improvement:
Demonstration of resource use or cost problems and opportunity for improvement, i.e., data demonstrating considerable variation cost or resource across providers

---

1a. High Impact or high resource use.

- This 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.
- The developer suggests that "by measuring population based relative resource use, health plans and providers can improve the affordability of health care without sacrificing quality."
- This measure was last endorsed in 2017, in which the developer cited 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.

1b. Opportunity for Improvement:

- The developer's primary care network (Minnesota, western Wisconsin, northeast Wisconsin, Iowa, North Dakota, South Dakota) consists of 65 individual provider groups that have 1,200 clinic sites. Provider group sizes vary from 600 to a few large systems with 50,000+ members.
- The developer states that performance was measured on an index basis relative to 1.00, where each one-point (0.01) variation from 1.00 (average) represents a 1% deviation from average.
    - The tables in 2b4.2 of the testing attachment show performance scores across measured provider groups highlighting variation of total resource use (2017 – 2019 dates of service).
- The developer expressed that meaningful difference in performance will vary by use of the measures due to differences in threshold. For example, a 5% increase in a providers' year over year performance would indicate a significant change, whereas a provider performing 5% higher than average might be considered average when developing a tiered network benefit.
- Out of the 65 provider groups measured in Total Resource Use measure:

    - 26 were better than average
    - 3 were 10% better than average
    - 12 were 10% higher than average
    - 50 were within 10% of the average
- Using Census tract data, the developer found a 1% increase in income resulted in a $0.13 decrease in total reimbursement, a $0.16 increase in resource use, and $0.28 decrease in price.
- The developers also found that insurance product also contributed with a $133 dollar difference in cost between commercial and Medicaid. The variation in resource use was much less, however, still significant with Medicaid covered members utilizing $75 more dollars of resources.

Questions for the Committee:

- *Has the developer demonstrated this is high impact, high-resource use area to measure?*
- *Is there a sufficient variation in performance across hospitals that warrants a national performance measure?*

Staff preliminary rating for opportunity for improvement: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:

Criteria 1: Importance to Measure and Report (including 1a, 1b)

1a. High Impact or High Resource Use: Has the developer adequately demonstrated that the measure focus addresses a high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality)?

- Yes, this maintenance measure of frequency and intensity of healthcare services from care providers at the population level can potentially address the key issue of high healthcare costs in the US.
- Yes
- Yes
- Yes, health care spending continues to be a significant proportion of the US GDP and variation in resource use is important in understanding value.
- Since endorsement in 2012, uptake of Total Resource Use measure has expanded across 42 states and is used by both regional and national organiztions. used to meet triple aim goals, optimizing health and patient experience while improving affordability. Last endorsed in 2017 and developer cited data demonstrating healthcare spending constitutes a high proportion of US GDP and high healthcare costs contributees to adults foregoin healthcare.
- Confirm importance
- No concerns
- Yes

1b. Opportunity for improvement: Was current performance data on the measure provided? Has the developer demonstrated there is a resource use or cost problem and opportunity for improvement, i.e., data demonstrating, considerable variation in cost or resource use across providers?

- The performance variation in provider groups used in the developer's sample (65 provider groups) seems only moderate. As a maintenance measure with almost 10 years of history, I would have appreciated if the measure developer included information about the historical adoption rate of the measure by new providers, and how their performance varied over time. It would have been informative if the developer could have incorporated any non-HealthPartners providers in the testing data.
- Yes
- Current data was provided. No considerable variation in cost.
- Yes, current data was provided reflecting 28.5% of providers were better than average, 16.5% were 10% better/higher than average, and 54.9% of providers were within 10% of the average. This provides evidence variation in resource use exists. The developers also found differences in cost and resource use between insurance products (Medicaid compared to commercial).
- Developer expressed meaningful difference in performance. Of the 65 provider groups measured, Total Resource Use measure showed 26 were better than average, 3 were 10% better than average, 12 were 10% higher than average and 50 were within 10% of the average. Other meaningful differences were noted.
- Opportunity for improvement depends on whether the measure truly represents resource use. See validity below
- No concerns
- Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Alignment of Specifications with Intent (includes threats to validity [e.g., attribution, costing method, missing data]) Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Multiple Data Sources; and Disparities.

Measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Measure evaluated by Technical Expert Panel? ☐ **Yes** ☒ **No**

Evaluators: N/A

Reliability

2a1. Specifications:

The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. All measures that use the ICD classification system must use ICD-10-CM.

2a2. Reliability testing:

Demonstration that the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

2a1 -2a2. Reliability:

- This measure was evaluated by the NQF Scientific Methods Panel (SMP). The SMP passed this measure on reliability with a High rating: H-4; M-3; L-0; I-2.
- Reliability testing was conducted at the performance measure score level:
    - The developer used two methods to demonstrate the repeatability of the results, using bootstrapped averages with full replacement and a 90 percent random sampling without replacement approach, which approximates controlling for case mix.
    - The variances from Actual RUI ranged from -0.0037 to 0.0062 in the bootstrap to -0.0019 to 0.0016 in the 90 percent sample.
- One SMP member commented that both the testing methods, bootstrap and 90% random sampling, theoretically work for large sample. However, it is unclear how the results change when dealing with smaller providers (providers with less than 600 members).
- Some other SMP members expressed that they would also like to see a signal-to-noise analysis.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *Do you have any concerns with the reliability testing that was not identified by the Scientific Methods Panel?*

Staff Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2a: Reliability

2a1. Reliability-Specifications: Describe any additional concerns you have with the reliability of the specifications that were not raised by the Scientific Methods Panel: Describe any data elements that are not clearly defined: Describe any missing codes or descriptors: Describe any elements of the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) that are not clear: Describe any concerns you have about the likelihood that this measure can be consistently implemented:

- None other than what the Scientific Methods Panel raised.
- tested only on Health Partners data set which may not be extensible
- None
- No additional concerns with the reliability of the specifications or that the measure can be consistently implemented.
- Noted that work for large samples but unclear how results chane wtih smaller providers (less than 600 members)
- No concerns
- No concerns
- No concerns

2a2. Reliability-Testing: Has the developer demonstrated that the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers? Describe any additional concerns you have with the reliability testing results or approach that were not raised by the Scientific Methods Panel.

- None.
- Reliable for Health Partners
- Measure is demonstrably reliable, when looking at large sample, but as noted it is unclear how this will perform with a smaller cohort and this will need to be adressed.
- The developer has demonstrated that the measure data elements are repeatable, producing the same results a high proportion of the time and precise enough to distinguish differences in performance.
- High reliability
- No concerns
- No concerns
- No concerns

Validity

2b1. Specifications align with measure intent:
The measure specifications are consistent with the measure intent and captures the most inclusive target population.
2b2. Validity Testing:
Demonstration that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided.
2b3. Exclusions:
Exclusions are supported by the clinical evidence**,** AND/OR There is a rationale or analysis demonstrating that the measure results are sufficiently distorted due to the magnitude and/or frequency of then on-clinical exclusions; AND Measure specifications for scoring include computing exclusions so that the effect on the

measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion); AND If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b4. Risk Adjustment:

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

2b5. Meaningful Differences:

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

2b6. Multiple Data Sources:

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

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

- This measure was evaluated by the SMP. The SMP Subgroup members found the measure to be valid and passed the measure with a generally High rating on validity (H-4; M-2; L-1; I-2;).

2b1. Specifications Align with Measure Intent:

- Attribution:
    - This measure is attributed to clinician groups and practices. This attribution approach was developed to fairly identify the provider that is responsible for managing the patient's care using the most recent visit as a tie breaker in a plurality model.
    - Members are also included in the measure by population or geographic location and assignment to a provider group regardless of where the patient received care.
- Costing approach
    - The risk adjusted Resource Use Index (RUI) measures the frequency and intensity of services utilized to manage a provider group's patients to give provider groups valuable information on resource use.
    - The measure score is a clinician's risk-adjusted paid per member per month basis as well benchmarked to a peer group.
    - This measure uses standardized pricing, the Total Care Relative Resource Values (TCRRVs), to payment standardize the claims.
        - TCRRVs are a grand linear scale of relative values, independent of price, designed to evaluate resource use across all types of medical services, procedures and places of service.
        - 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.

2b2. Validity Testing:

- Three approaches were used to assess validity:
    - critical data elements were correlated with each other and utilization

    - performance measure score was validated against “known risk-adjusted utilization metrics”
    - high/low performing groups were compared on utilization and RUI (although this largely shows the validity of the risk adjustor), and face validity was determined through a 45-day comment period.
- Validity of Measure Components
    - There is a high correlation between ACG score and the non-risk adjusted PMPM and Non-Risk Adjusted Total Cost Relative Resource Values (TCRRVs), which indicates that the non-risk adjusted PMPM and the non-risk adjusted TCRRVs are a good measure of resource use.
    - Correlations between the Non-Risk Adjusted Place of Service Metrics and Non-Risk Adjusted PMPMs & Non-Risk Adjusted TCRRVs are strong (ranging between 0.55 and 0.84).
    - The non-risk adjusted resource composite is highly correlated with ACGs, non-risk adjusted PMPMs, and non-risk adjusted TCRRVs (ranging between 0.77 and 0.95).
    - Professional RUI is correlated with overall RUI.
- Validity of measure score:
    - The indexed Total Resource Use measure has a high correlation to a risk-adjusted composite utilization index, which was developed as a proxy to measure total resource consumption.
- An assessment of high and low performing provider groups supports the relationship between risk adjusted utilization metrics and RUI.
- SMP reviewers generally agreed that the empirical validity tests are appropriate and in the expected direction. However, one SMP member expressed interest in seeing the relationship of this measure to a quality construct.
- A few SMP members commented on the use of ACG scores, stating that using a proprietary model limits the transparency and generalizability.

2b3. Clinical Inclusions and Exclusions/Evidence to Support Clinical Logic

- The HealthPartners’ Total Resource Use measure has the following exclusions and decision points:
    - Nine-month continuous enrollment – A nine-month continuous enrollment was selected to balance business operations. Nine months allows for partial year enrollee. There was very little statistical difference in R-squared between six and twelve months.
    - Infants, under the age of one, are excluded due to slightly higher R-squared of the population without newborns, the required nine months enrollment criteria and variability in newborn costs, newborns under the age of one were excluded from the total cost of care measure.
    - Members over age 64 due are excluded due to potential incomplete claims data of Medicare eligible beneficiary.
    - The truncation level of 125,000 TCRRVs is used for the measure as the percent of members truncated has been stable over the last few years (0.3%-0.4%). Truncation levels are reviewed to stay aligned with healthcare costs and ensure stability of the measure. TCRRVs per member above 125,000 are excluded (truncated).
    - The result of these exclusions is 77.6% are maintained in the measure population.
- The measure includes:
    - Administrative claims covering all categories of health care services: professional, facility inpatient and outpatient, pharmacy, lab, radiology and any other ancillary healthcare services
    - Johns Hopkins ACG System for risk adjustment
    - Membership eligibility, identifier and number of months during the measurement period the member was eligible (member months)

2b4. Risk adjustment

- The Total Resource Use measure uses the proprietary 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 assigns International Classification of Disease (ICD) diagnosis codes to 32 diagnosis groups – Aggregated Diagnosis Groups (ADGs). 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.
- ACGs are defined by morbidity, age, and sex. Based on the pattern of morbidities, the ACG approach assigns each individual to a single ACG category.
    - Each unique member is assigned one of 93 ACG actuarial cells, which has a corresponding weight that reflects relative illness burden (e.g., relative expected resource consumption). Attributed members are assigned a risk score based on diagnoses on claims from the performance measurement period, as well as member age and gender.
    - Each actuarial cell can be considered a covariate of the multivariate risk model with the cell weights being the coefficients.
- The ACG score is determined when each provider's attributed member ACG weights are summed to the provider level and divided by the sum of the attributed member months creating an ACG score for the provider.
    - The provider's average ACG score is indexed to all attributed member's plan average ACG score.
    - A member's total member ACG weight is updated to correspond with each year the Total Resource Use measure is measured.
- The developer considered education and income for risk adjustment, but did not include these social risk factors in the final model.
- The developer reports a range of R-squared values:

Metric	R-Squared: ACG	R-Squared: Non-Risk Adj PMPMs
Non-Risk Adj PMPM	0.45	1.00
Non-Risk Adj TCRRVs	0.90	0.55
ACG Risk Adj TCI	0.00	0.52
ACG Risk Adj RUI	0.01	0.07
Price	0.00	0.36

- Some SMP reviewers questioned the lack of race and social risk factors, especially about how income as a social risk factor was considered in the risk adjustment model.

2b5: Meaningful Differences

- The developer states that the Total Resource Use measure can effectively identify variation in performance levels.

- Out of the 65 provider groups measured in Total Resource Use measure, the scoring reflected:
    - 26 were better than average
    - 3 were 10% better than average
    - 12 were 10% higher than average
    - 50 were within 10% of the average

2b6. Multiple Data Sources

- The developer did not specify multiple data sources.

2c. Disparities

- Using Census tract data, the developer found a 1% increase in income resulted in a $0.13 decrease in total reimbursement, a $0.16 increase in resource use, and $0.28 decrease in price.

  The developers also found that insurance product also contributed with a $133 dollar difference in cost between commercial and Medicaid. The variation in resource use was much less, however, still significant with Medicaid covered members utilizing $75 more dollars of resources.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., correlations, exclusions, risk-adjustment approach, etc.)?*
- *Does the SC have any concerns related to the risk adjustment model (e.g., the r-squared values, lack of social risk factor adjustment)*

Staff preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2b: Validity

2b1. Validity -Testing: Describe any concerns you have with the testing approach, results and/or the Scientific Methods Panel's evaluation of validity: Describe any concerns you have with the consistency of the measure specifications with the measure intent: Describe any concerns regarding the inclusiveness of the target population: Describe any concerns you have with the validity testing results: Does the testing adequately demonstrate that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided?

- None other than those raised in the methods panel.
- yes
- Testing is adequate
- No concerns with the testing approach or the SMP's evaluation of validity. There is a high correlation with the index and known utilization metrics
- High correlation however a propriety model may limit transparency and generalizability. Not clear on the specifications
- I question whether claims data represents cost or resource use. Cost to whom? Insurance company, facilities, clinicians, patients, caregivers. Claims data reflects $ paid by insurer according to benefits, no relation to cost. Resources include facility, treatment, studies, human, and out-of-pocket and life. Doesn't include staff turnover, training, burnout or out-of-pocket co-pays, transportation, dependent care, loss of wages, etc.
- No concerns
- No concerns

2b2. Additional threats to validity: Describe any concerns of threats to validity related to attribution, the costing approach, or truncation (approach to outliers): Attribution: Does the accountable entity have reasonable control over the costs/resources measured? Is this approach aspirational (intending to drive change) or was it developed based on current state? Costing Approach: Do the cost categories selected align with the measure intent, target population and care settings? Is the approach for assigning dollars to resources agreeable? Truncation (approach to outliers): What is the threshold for outliers (i.e., extremely high cost or low cost cases) and are they handled appropriately?

- No concern with attribution and the costing approach. However, the truncation of members' cost at $125K is somewhat arbitrary. Why $125K but not $100K.
- yes, though the process for imputing pharmacy costs is concerning
- Attribution - Approach is aspirational. Costing - Aligned & assignment agreeable, my only concerns are that by excluding members over age 64 we limit disproportionately the groups that likely are sicker and utilize care at a greater level than those under 64. Truncation -
- No concerns regarding the attribution or costing approach. Inclusion/exclusion includes excluding TCRRVs per member above 125,000. This criteria is described as a "truncation level." Would like to know more about the rationale for exclusion rather than capping at a lower value. Are there certain providers which this may affect more?
- Not clear if SES, race/ethnicity or SDoH are incorporated
- Don't understand exclusion of >64 Medicare data due to incomplete Medicare claims data. Does NQF never use Medicare claims data in measures? Why is it an issue here?
- No concerns
- No concerns

2b3. Additional Threats to Validity: Exclusions Describe any concerns with the consistency exclusions with the measure intent and target population: Describe any concerns with inappropriate exclusion of any patients or patient groups:

- None
- none
- See note above, I am not convinced that excluding patients over 64 because the Medicare data is delayed is an appropriate rationale.
- This comment is also included in the truncation section since it is described as truncation in the Measure Worksheet. Inclusion/exclusion includes excluding TCRRVs per member above 125,000. This criteria is described as a "truncation level." Would like to know more about the rationale for exclusion rather than capping at a lower value. Are there certain providers which this may affect more?
- NA
- none
- No concerns
- High rating for validity

2b4/2c. Additional Threats to Validity: Risk Adjustment Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factors that were available and analyzed align with the conceptual description provided? Has the developer adequately described their rationale for adjusting or stratifying for social risk factors? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Describe any concerns with the appropriateness of risk adjustment (case-mix adjustment) development and testing: Do analyses indicate acceptable results?

- 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). Income and education were also tested but not readily available, particularly, for commercially insured patients. I thought the developer's rationale for testing only for income was acceptable. Results from the testing appears acceptable, and no additional concerns.
- I think we need a different approach for looking at social risk factors because the testing data do not match the reality that race and income do impact outcomes. Testing the addition of just one or two SES factors obscures the impact due to partial effects for other variables in the model (e.g. comorbidities). Stratifying results by SES would likely reveal more disparities, and testing SES in a different way.
- I believe that limiting to commercial payors only will/has excluded a large cohort that would have provided informative data to this program and has a negative impact on the appropriateness of the case mix
- I agree with the SMP questions regarding the lack of race and social risk factors in risk adjustment. The developers explored income as a risk adjustment variable and ultimately did not include it in the model. In terms of disparities, the developer describes "Using Census tract data, the developer found a 1% increase in income resulted in a $0.13 decrease in total reimbursement, a $0.16 increase in resource use, and $0.28 decrease in price." Additionally, the developer found differences in cost and resource use for commercial versus Medicaid covered members. Given the relationship between Medicaid eligibility and income, would like to know more about the rationale for excluding income from the final model.
- Yes
- DK
- No concerns
- Yes

2b5. Threats to Validity: Meaningful Differences Describe any concerns with the analyses demonstrating meaningful differences among accountable units:

- None
- none
- None
- No concerns.
- NA
- DK
- No concerns
- No concerns

2b6. Threats to Validity: Missing Data/Carve Outs - Describe any concerns you have with missing data that constitute a threat to the validity of this measure: Carve Outs: Has the developer adequately addressed how carve outs in the data source are handled (or should be handled for other users)? For example, if pharmacy data is carved out (missing) from the data set, can a measure that focuses on cost of care the target clinical population still be valid?

- None
- Imputing pharmacy data is probably not the answer for other organizations to use this data. Testing on an organization with a homogenous data set may not reflect reality for other organizations.
- No concerns
- No concerns.
- NA
- DK
- No concerns
- No concerns

Criterion 3. Feasibility

3. Feasibility

The extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer states that data elements are generated from health plan claims.
- The developer shares that all data elements are in defined fields in a combination of electronic sources
- After receiving general feedback, the developer made the following changes were made to expand use:
    - HealthPartners organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download necessary tools to run the measure, free of charge. In addition,
    - HealthPartners organized non-SAS user instructions for organizations without the software.
- The measure and software are available free of charge at www.healthpartners.com/tcoc.
- The developer states that the ACG System is widely available within the public and private sectors in the US and abroad.

- The pricing for the ACG System varies for commercial and government entities but is generally based on a per member per year license that is tiered to provide lower per member costs for larger entities. For a commercial plan there is a base fee of $33,000 annually with incremental costs between $0.05 and $0.41 per member per year based on volume, which is inclusive of both license fees and support.
- Discounted fees are available for government entities and other grant funded not-for-profit entities.
- Additionally, Johns Hopkins offers research licenses for academic users incorporating ACGs into published research.

Questions for the Committee:

- *Are there any concerns regarding feasibility?*

Staff preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? Describe your concerns about how the data collection strategy can be put into operational use: Describe any barriers to implementation such as data source/availability, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary tools (e.g., risk adjuster or grouper instrument):

- Data elements are commonly and routinely generated in the EHR systems. While the measure calculation seems complicated, HealthPartners provide the necessary tools (SAS programs etc.) free of cost to providers willing the compute this measure for their own data.
- ACG is proprietary
- It appears that barriers to access/implementation have been addressed by providing training and software for free to organizations, and by collecting data from the plan claims remove the issues of EHR utilization..
- The measure seems feasible to implement since it uses claims data and widely available software to calculate the ACG Score
- From health claims
- no comment
- Implementation requires the use of ACGs which must be licensed separately and may be cost-prohibitive for some entities.
- Moderate rating for feasibility

Criterion 4: Usability and Use

Use

4a. Use. evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency.

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

4a.2. Feedback on the measure by those being measured or others.

Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

---

4a1. Current uses of the measure

- **Publicly reported? ☒ Yes ☐ No**
- **Current use in an accountability program? ☒ Yes ☐ No ☐ UNCLEAR**

Accountability program details

- The developer states that this measure has been used in accountability applications and publicly reported in 42 states in the country and used by both national and regional organizations for driving improvement, since endorsement in 2012.
- This measure is used in Public Reporting, Payment Program, and Quality Improvement with Benchmarking at:
    - HealthPartners
        - As a health plan, HealthPartners uses the Total Resource Use measure to incentivize providers to meet Triple Aim goals, optimizing health and patient experience while improving affordability.
        - HealthPartners publicly reports provider group cost results for purposes of transparency for employers, providers, and consumers. The resource use results are paired with Total Cost of Care, quality and experience metrics to promote quality improvement with benchmarking across providers.
        - HealthPartners has shared savings payment agreements with over 85% of its primary care providers which has increased provider engagement and sharing of appropriate risk as a partnership to lower cost for providers and patients while maintaining quality and experience.
        - HealthPartners has begun building upon it by implementing new payment reform models that align incentives among specialists and hospitals to support shared savings with primary care.
    - Network for Regional Healthcare Improvement
        - Eleven regions were part of a project to develop a standardized method of reporting total cost and resource use by using the HealthPartners endorsed measures.
        - During 2015, seven regions were able to share healthcare cost information on over five million patients attributed to 20,000 individual physicians through practice level reports.
- This measure is used in Public Reporting and Quality Improvement with Benchmarking at Minnesota Community Measurement.
    - Beginning in 2016, MNCM publicly reported Total Resource Use data by provider group in Minnesota using HealthPartners endorsed Total Resource Use measure.

    - Through their multi-stakeholder collaborative process, they were able to collect cost data from four health plans and publicly spread the use of the measure to all provider groups in Minnesota.
- This measure is used in the following payment programs: The Alliance, Public Health/Disease Surveillance, The University of Iowa Public Policy Center, Delaware Health Information Network (third party of Delaware state agency).
- This measure is used for quality improvement with benchmarking at the following regional collaboratives: Maine Health Management Coalition, Oregon Health Care Quality Corporation, HealthInsight and Utah Department of Health, Washington Health Alliance, Center for Improving Value in Health Care (CIVHC) and Midwest Health Initiative.
- This measure is used for quality improvement at the following Priority Health, Providence Health Plans, Onpoint Health Data and Provider Groups in Minnesota, Wisconsin, and Iowa.

4a2. Feedback on the measure by those being measured or others

- The developer provides quarterly comprehensive reports and monthly patient applications to best support providers in identifying opportunities for improving affordability for their patients, while at the same time supporting patient outreach, pre-visit planning, and care coordination efforts.
- The developer engages with its provider groups and network on an ongoing basis and also organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download the necessary tools to run the measure, free of charge.
    - HealthPartners has created instructions and toolkits for both SAS and non-SAS users.
- The developer provided specific examples of usability and quality improvement efforts shared by organizations locally and nationally under the 'Quality Improvement' section in the following link: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf
- The developer has addressed feedback by:
    - revising the truncation limit criteria based on rising health care cost.
    - continuously addressing questions and clarifying details to support measure implementation and review.
    - preparing user guides and maintaining public-facing website which includes several resources and technical documentation.

Additional Feedback:

- N/A

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

Staff preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

Usability

4b. Usability.

The extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement.

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

4b2. Benefits vs. harms.

Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b1. Improvement results

- The developer stated that there was great improvement from this measure.
- One specific is example is the Northwest Metro Alliance, which serves more than 300,000 people receiving care at nine different clinics and one hospital along with its affiliated specialists. From 2010 to 2019, the alliance achieved medical cost trends that were 6% lower than the Minnesota Twin Cities metro trend percentage. This decrease has resulted in $40 million in savings when compounding the Northwest Metro Alliance trend over 10 years compared to the Twin Cities metro.
- The developer reported that there were cases in which some providers' business model is based on revenue generation and in these cases, they did not see total cost or resource use improvement which also signals effectiveness of the measure's ability at indicating performance.

4b2. Unintended consequences

- The developer did not identify any unexpected findings.

4b2. Benefits vs. harms

- The developer did not identify any unexpected benefits or potential harms.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

Staff preliminary rating for Usability and Use: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Is the measure being used in any other accountability applications? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Is a credible plan for implementation provided?

- How is the measure being publicly reported? The measurement document states that since endorsement in 2012, uptake of the Total Resource Use measure has expanded across 42 states in the country and used by both national and regional organizations (Coverage) and they provided in a weblink to the find more details. However, I was expecting to see the names of the organizations that adopted this measure by states, which I could not find in the document that the weblink pointed to. Is the measure being used in any other accountability applications? The measure has been used in accountability applications and publicly reported in multiple states for driving improvement. However, as far as I can tell, these multiple states include only the states with HealthPartners providers. In addition, these accountability applications include only public reporting and for informing payment agreements with HealthPartners (as opposed to being used across the board). Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Beginning in 2016, Minnesota Community Measurement publicly reports Total Resource Use data by provider group in Minnesota using HealthPartners endorsed Total Resource Use measure. However, I am not sure if similar initiatives are in place in other states that uses Total Resource Use Data. Is a credible plan for implementation provided? It is reported to be "not applicable" in the measure document.
- specifications not fully transparent
- No concerns
- The developer states that this measure has been used in accountability applications and publicly reported in 42 states in the country and used by both national and regional organizations for driving improvement, since endorsement in 2012.
- Publically reported in 42 states
- yes
- n/a
- Yes

4a2. Use - Feedback on the measure: Describe any concerns with the feedback received or how it was adjudicated by the measure developer: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data?Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation?Has this feedback has been considered when changes are incorporated into the measure?

- Use - Feedback on the measure: Describe any concerns with the feedback received or how it was adjudicated by the measure developer: The developer states that it continues to be engaged with provider groups on an ongoing basis throughout the year and have a partnership relationship with our provider network. The provide the following linke to see the feedback etc. However, I did not find any feedback from measured entities nor did I find how such feedback were addressed: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Yes, the measured entities are given performance results or data, as well as assistance with interpreting the measure results and data. Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Yes. It appears that HealthPartners work with the providers being measured. However, documentation on those feedback and their responses were not available. Has this feedback has been considered when changes are incorporated into the measure? It appears that the measure developer (HealthPartner) takes into account the input/feeback from the measured entities. As an example, it states that it revised the truncation limit criteria over the years based on rising health care cost. It also holds education sessions, and continuously address questions and clarify detail to support measure implementation and review.
- after all these years of using the measure, I was disappointed not to see more data showing the correlation of this measure to quality, yet the assumption that less is always better.
- Developer demonstrates that they address the feedback provided.
- No concerns with feedback. The developer provides quarterly reports and applications to support providers in using the measure as well as stakeholder engagement opportunities, and free resources on their public facing website.
- Yes
- No comment
- Used primarily by the developer
- Yes

4b1. Usability – Improvement: Has the measure developer demonstrated that the use of this measure is helping to drive improvements in cost or efficiency? Has the developer adequately described how the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations?

- Has the measure developer demonstrated that the use of this measure is helping to drive improvements in cost or efficiency? Providers groups, both local and national, that used this measure since its endorsement in 2012 seems to have improved transparency, efficiency and cost of care, resulting in significant shared savings. Has the developer adequately described how the performance results be used to further the goal of high-quality, efficient healthcare? Yes. If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? N/A
- did not provide data that shows care is high-quality, just that resource use declined
- No concerns
- The developer provides an example of one organization using the measure across multiple clinics achieving significant savings and lower costs.
- No issues

- I can't see how this measure informs or drives improvement in resource use benefiting patients and caregivers or direct care clinicians for that matter. How has this improved anything since begun in 2012, nine years ago. Has resource used changed, for whom, where?
- Developer displays some improvement data and seems to suggest causation but without adequate comparisons this is difficult to conclude.
- Yes

4b2. Usability – Benefits vs. harms: Describe any unintended consequences and note how you think the benefits of the measure outweigh them:

- None.
- none
- None noted
- None.
- Benefits outweight harm
- No comment
- No concerns
- No identified concerns

Criterion 5: Related and Competing Measures

- The developers stated that there were no related or competing measures for this measure.

Harmonization

- N/A

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- None.
- not clear why there are no related measures.
- None noted
- There are no related to competing measures.
- None
- DK
- The developer states there are no competing or related measures but what about their own total cost of care measure for the same populations?
- None

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/17/2021

- **No NQF members have submitted support/non-support choices as of this date.**

Combined Scientific Methods Panel Preliminary Analysis of Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 1598

Measure Title: Total Resource Use Population-based PMPM Index

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☒ No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member 1: (MIF: S.9.1) "TCRRVs per member up to 125,000 are included; TCRRVs per member above 125,000 are excluded (truncated)" – what was the basis for choosing the max threshold of 125,000 of TCRRVs per member? (MIF: S.9.1) Has the measure developer tried what would have happened if they were to restrict the measurement sample to only those with both medical and pharmacy benefit? I am just worried that the people without pharmacy benefit have different risk profile, and thus they might have introduced selection bias by including those patients with pharmacy benefit. (MIF: S.9.4) "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. 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." – I am still not clear how the actual values were created to begin with? (MIF: S.13.2) In terms assignment of a member to a provider, the measure developer states that "Assignment – each member is assigned to a responsible provider group regardless of where he or she received care". This seems odd that patients will be assigned to a provider group regardless of where s/he received care? Then what's the point of assignment?

   Panel Member 2: None

   Panel Member 4: Risk adjustment uses the ACG, a proprietary algorithm

   Panel Member 5: NA

   Panel Member 6: Measure includes a proprietary risk adjustment system (ACGs) and cannot be replicated by the measured organizations. Transparency is an important component in measurement and improvement activities.

   Panel Member 7: If this is a "population" measure—and the population is all patients in HealthPartners, what is the comparison group—patients not in HealthPartners? Calculation methodology of "a risk-adjusted paid per member month basis as well as benchmarked to a peer group" requires operational specification. Also, the measure description states that the measure "is a risk adjusted measure of the frequency and intensity of services" as compared with the Total Cost of Care (NQF #1604) that focuses on health care costs. These two statements seem contradictory. NOTE: none of the articles regarding measure performance (IM1.3) is more recent than 2016—before the most recent re-endorsement.

   Panel Member 8: No concerns.

RELIABILITY: TESTING

Type of measure:

☒ **Outcome (including PRO-PM)** ☐ **Intermediate Clinical Outcome** ☐ **Process**

☐ **Structure** ☐ **Composite** ☒ **Cost/Resource Use** ☐ **Efficiency**

Data Source:

☐ **Abstracted from Paper Records** ☒ **Claims** ☐ **Registry**
☐ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs**
☐ **Instrument-Based Data** ☐ **Enrollment Data** ☐ **Other (please specify)**

Level of Analysis:

☐ **Individual Clinician** ☒ **Group/Practice** ☐ **Hospital/Facility/Agency** ☐ **Health Plan**
☒ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☒ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   ☒ **Yes** ☒ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member 1: To measure the reliability of the RUI measure the actual results were compared to the results calculated by two sampling methods, bootstrapping and a 90% random sample. The mean Total Resource Use (RUI) from these 500 iterations was computed and compared to the Actual Resource Use Index (RUI) index for each provider group. The mean of the sampled Total Resource Use index was calculated for each provider group and compared to the Actual RUI index for each provider group.

   Panel Member 2: Method is basically sound but would like more statistics on how distribution/classification changes across the bootstrapped samples.

   Panel Member 4: The developer used two methods to demonstrate the repeatability of the results, using bootstrapped averages and a 90% random sampling replacement approach which approximates controlling for case mix. These methods are novel and appear appropriate.

   Panel Member 5: The variances from Actual RUI ranged from -0.0037 to 0.0062 in the bootstrap to -0.0019 to 0.0016 in the 90% sample. The mean Total Resource Use results from the bootstrap and 90% samples compared to the actual RUI results for each provider group are displayed on the charts on the following pages. The variance between the actual RUI to the bootstrap results is shown on the far right of each chart. The RUI charts are sorted in ascending order by Total Cost Index. See Reliability Paper for detailed results.

Panel Member 6: Bootstrap and random sample method - both similar. Performed at the score level.

Panel Member 7: The documentation states that this is a population measure and hence does not require confidence intervals. However, in their S.13.4 Sample size narrative they state that "This measure has been tested for a minimum attributed member population of 600 members…" indicating that the measure is NOT a population measure, but a subpopulation measure and is used for comparison of these subpopulations across the total population of HealthPartners. A 90% sample of the population should produce essentially identical results to the population, and selecting and then averaging 500 sample results (their bootstrap methodology) should produce the same. These methodologies are not convincing.

Panel Member 8: Bootstrapping patient samples with and without replacements is a clever way to assess reproducibility of the measure results.

Panel Member 9: To measure the reliability of the RUI measure the actual results were compared to the results calculated by two sampling methods, bootstrapping and a 90% random sample. In the 90% random sample method, the members that were attributed to a provider group were randomly sampled at the 90% membership level without replacement. This technique was employed to create variation within a provider group by leveraging their own population and controlling for the patient case mix variation that is introduced when random sampling is employed.

7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   Panel Member 1: The variances from Actual RUI ranged from -0.0037 to 0.0062 in the bootstrap to -0.0019 to 0.0016 in the 90% sample. Thus, both the bootstrap and 90% random sampling methods closely mirrors the actual index.

   Panel Member 2: Minimum patient panel size of 600 produces narrow variation around mean scores. Given panel size, reliability based on typical variation within groups is small relative to cross-group variation. Would have liked a more formal set of statistics to complement the analysis done.

   Panel Member 4: The bootstrap and random replacement approaches showed very little difference in within provider variation.

   Panel Member 5: The results of the Bootstrap and Random Sample tests allow us to confidently conclude that the measures will reliably decipher RUI performance between levels of analysis (e.g., provider group). The bootstrap results indicate that the RUIs are reliable as the provider variation within all groups is <1% whereas the variation between groups spans >110%.

   Panel Member 6: Results presented in a table with only range RUI is reported. Difficult to assess. There is a rather long table listing variance by provider. There does not appear to be a pattern (good), but not the best presentation of the results.

   Panel Member 7: As expected, there were few differences between population and "provider level" results.

   Panel Member 8: Strong within practice stability in scores.

   Panel Member 9: The variances from Actual RUI were small, ranging from -0.0037 to 0.0062 in the bootstrap to -0.0019 to 0.0016 in the 90% sample. The results of the bootstrap and random sample tests show that the measures will reliably decipher RUI performance between levels of analysis (e.g., provider group). The bootstrap results indicate that the RUIs are reliable as the provider variation within all groups is <1% whereas the variation between groups spans >110%.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☒ **No**

☒ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☒ **No**

   ☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and **all** testing results):

    ☒ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has **not** been conducted)

    ☐ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Panel Member 1: Both the testing methods, bootstrap and 90% random sampling, theoretically work for large sample. I am not sure how the results change when dealing with smaller providers (providers with less than 600 members). The measure developer says in the MIF document that that additional reliability and validity checks are needed without clearly delineating what those checks would be. Thus, for smaller providers, the reliability of the measure would be in question.

    Panel Member 2: Random sampling to estimate within group variance is reasonable but would like to have seen a more formal application of a S/N type measure.

    Panel Member 3: There is no reason why the developer cannot use an ICC or signal to noise analysis (like every other measure in this category)

    Panel Member 4: The developers used a compelling method to show that within provider RUI variation was very small.

    Panel Member 5: o The approach demonstrated consistency in the performance score across the specified samples.

    Panel Member 6: Critical data element testing was not conducted. I believe this is OK according to the decision tree. If not, please let me know and I will reassess.

    Panel Member 7: I do not understand why NQF is endorsing a measure that is described as a "population" measure for one health care system (i.e., HealthPartners)—that they use to compare individual Provider entities within their system. If a measure is not intended for comparison across all health care systems (e.g., HCA, KP), why is NQF offering endorsement of an internal metric for one health care system (i.e., HealthPartners)?

    Panel Member 8: Strong conceptual method for assessing reliability; strong results.

VALIDITY: TESTING

12. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☒ **Both**

13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

NOTE that data element validation from the literature is acceptable.

Submission document*: Testing attachment, section 2b1.*

☒ **Yes**

☒ **No**

☒ **Not applicable** (data element testing was not performed)

14. **Method of establishing validity of the measure score:**

☒ **Face validity**

☒ **Empirical validity testing of the measure score**

☐ **N/A (score-level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☒ **No**

☐ **Not applicable** (score-level testing was not performed)

16. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

Panel Member 1: Face validity as well as empirical validity measures adopted for testing describes the hypothesized relationships well.

Panel Member 2: Correlations across adjusted and unadjusted measures and correlation of components appear reasonable. Predicted magnitudes and directions derived from expectations of performance of measure.

Panel Member 4: Three approaches were used to assess validity: critical data elements were correlated with each other and utilization, performance measure score was validated against 'known risk adjusted utilization metrics' and high/low performing groups were compared on utilization and RUI (although this largely shows the validity of the risk adjustor), and face validity was determined through a non-standardized 45 day comment period. I did not consider the face validity results since they were not performed in accordance with NQF standards.

Panel Member 5: Critical data elements Non-risk adjusted correlations between ACG and Total Resource Use, Total Cost Relative Resource Values (resource use) and utilization metrics were calculated. Performance Measure Score Risk adjusted Resource Use Index correlations to known risk adjusted utilization metrics were calculated. Empirical testing of validity and overview of face validity policy and procedure. An assessment of high and low performing provider groups supports the relationship between risk adjusted utilization metrics and Resource Use Index. The face validity process is conducted by transparently sharing results and methods with provider groups measured and allowing a 45-day comment period prior to public display of provider group results. o HealthPartners has a Policy and Procedure Review Process and executes it annually with each release of provider groups' performance and measurement results.

Panel Member 6: Appropriate - correlations with related measures.

Panel Member 7: The comparisons provided in the attached validity paper provided information about both "risk-adjusted" (Appendices A & B) and "non-risk adjusted composite utilization"(Appendix C)

whereas the measure is described as “risk-adjusted”. Hence, the information provided may not relate to the proposed measure.

Panel Member 8: The authors looked at correlation between resource use and several other measures - I like the hypothesis testing with expected positive and negative relationships between resource use and criterion.

Panel Member 9: At the data element level the developers evaluated non-risk adjusted correlations between the ACG and Total Resource Use, Total Cost Relative Resource Values (resource use) and utilization metrics. At the performance measure score level, risk adjusted Resource Use Index correlations to known risk adjusted utilization metrics were calculated. Empirical testing included an assessment of high and low performing provider groups that demonstrated the relationship between risk adjusted utilization metrics and Resource Use Index. The face validity process was conducted by transparently sharing results and methods with provider groups measured and allowing a 45-day comment period prior to public display of provider group results.

17. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

Panel Member 1: The face validity process is conducted by transparently sharing results and methods with provider groups measured and allowing a 45-day comment period prior to public display of provider group results. As demonstrated, the measure components correlations (in 2b1.3) are all in expected directions, as well moderate to high range. With regard to the Total Resource Use Index, the measure developer claims that the Total Resource Use is correlated with TCI as expected (2b1.3). However, as far as I can see, that correlation is only 0.19. Similarly, the developer’s claim that Professional RUI is correlated with overall RUI, supporting the notion primary care providers are integral in the management of total costs and resources is based on a correlation of 0.52. The rest of the correlations detailed in section 2b1.3 demonstrating validity appears to be in acceptable range.

Panel Member 2: Yes.

Panel Member 4: The developer used appropriate approaches to evaluate validity by comparing high and low performing groups utilization and RUI. Developer also looked at correlations between correlations between resource use and ACG grouper score.

Panel Member 5: In summary, the Total Resource Use measure accurately and consistently identified providers that are low or high performers with conclusions supported by known utilization measures. There are high correlations between non-risk adjusted PMPM, ACG score and non-risk adjusted TCRRVs which indicate they are good measures of resources. The ACGs, non-risk adjusted PMPMs, and non-risk adjusted TCRRVs have similar correlations to all utilization metrics which indicates the TCRRVs are performing as expected and are a solid measure of resources. The indexed Total Resource Use measure scores have a moderately high correlation (0.47) to a risk adjusted composite utilization index score, which was developed as a proxy to measure total resource consumption. The Total Resource Use measure differentiates between provider groups accurately as supported by the risk adjusted service utilization metrics.

Panel Member 6: The testing was only conducted with the submitters provider groups - not clear if this is generalizable to other providers.

Panel Member 7: Appendix A provides a 3-year display of RUI results. However, there is no clear explanation (or analysis) on why changes occurred year-over-year for the individual Providers, nor if there was a pattern between RUI changes and TCI or Price Index changes across these same years that would establish concurrent validity with the RUI measure. NOTE: Section 2b1.3 of the Testing document does provide some results that relate to demonstrating validity of the RUI measure, although even these

presentations reference “non-risk adjusted” results. Additionally, several risk-adjusted RUI correlations with other measures are < 0.50 that would indicate low correlation.

Panel Member 8: Lots of results - overall positive picture.

Panel Member 9: There was good correlation between the ACG Score and the Non-Risk Adjusted Per Member Per Month (PMPMs) (0.67), strong correlation between the ACG and Non-Risk Adjusted Total Cost Relative Resource Values (TCRRVs) (0.95) and between the Non-Risk Adjusted Per Member Per Month (PMPMs) (0.74). There was low correlation between the ACG risk adjusted Total cost index score and ACG (0.11) and with the non-risk adjusted PMPMs (0.26). This indicates that the risk score of a provider has no impact on a provider’s ability to be a high performer. There is a low correlation between price and ACG (-0.06) because ACGs measure expected resource use whereas price is not affected by the number or intensity of services received. Correlations at the service level showed similar performance. Correlation between the non-risk adjusted PMPM for inpatient was 0.75, outpatient ER 0.62, outpatient surgery 0.63, professional ranged from 0.53-0.70, and pharmacy 0.79. They also created a composite utilization metric by weighting each of the underlying utilization metrics across places of service by the percent of resources it represents. Correlations between the composite index with the ACG were 0.91, the non-risk adjustment PMPM 0.77, and with the non-risk adjusted total cost values 0.95. Total Resource Use is strongly correlated with total cost index as expected (0.78). Professional resource use was less strongly correlated with the RUI (0.52) and hospital resource use less correlated (0.28) since a lower proportion of patients have a hospitalization. They also evaluated correlations between risk adjusted place of service utilization metrics and RUI: inpatient 0.73, outpatient ER 0.42, outpatient OP surgery 0.66, professional ranged from 0.21 to 0.63. The risk adjusted composite correlation with RUI was 0.47. They also evaluated the stability and sensitivity of the measure over time and found providers’ performance was relatively consistent across 3 years. In summary, the Total Resource Use measure seems to be able to consistently identify providers that are low or high performers with conclusions supported by known utilization measures. There are high correlations between non-risk adjusted PMPM, ACG score and non-risk adjusted TCRRVs which indicate they are good measures of resources. The ACGs, non-risk adjusted PMPMs, and non-risk adjusted TCRRVs have similar correlations to all utilization metrics which indicates the TCRRVs are performing as expected and are a solid measure of resources. The indexed Total Resource Use measure scores have a low/moderate correlation (0.47) to a risk adjusted composite utilization index score, which was developed as a proxy to measure total resource consumption.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

Submission document: Testing attachment, section 2b2.

Panel Member 1: Not clear to me why and how the threshold of $125,000 was chosen as to exclude members with more than that threshold.

Panel Member 2: No exclusions. Relies heavily on ACG risk adjuster to control for expected variations in resource use and cost.

Panel Member 4: No concerns

Panel Member 5: None

Panel Member 6: None.

Panel Member 7: There are exclusions for age. This may be problematic in the RUI measuring true “population” resource use as older clients (>65) tend to use different (more?) resources than younger clients.

Panel Member 8: None.

Panel Member 9: Exclusions are few and seem appropriate.

19. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

19a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification** ☒ **Other**

Panel Member 1: 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.

Panel Member 5: ACG risk grouper

Panel Member 6: Risk adjustment based on ACG scores.

Panel Member 7: Measure uses Johns Hopkins Adjusted Clinical Group actuarial methods.

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☒ No

19d. **Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☒ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☐ No

19d.5. Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

19e**. Assess the risk-adjustment approach**

Panel Member 1: In the supplementary link, the developer provide rationale for testing the inclusion of only income variable but not race and other SES indicators.

Panel Member 2: Use well documented and established Hopkins ACG model. Appears to include contemporaneous diagnoses, which may be appropriate for a total cost of care model.

Panel Member 4: Unable to know whether RFs are at start of care due to use of proprietary grouper, therefore, I am unable to answer questions about adequacy of the risk adjustment approach.

Panel Member 5: Used a reasonable method, but with continued disparities in health, there was limited consideration of how income is bidirectional with clinical factors and other health determinants.

Panel Member 6: Also concerned with the lack of assessment of the applicability of the ACG scores to their patient population. How do the demographics of the patients measured compare to those used to calibrate the ACG scores. Submitted assumes that the ACG scores are applicable to their population without assessment.

Panel Member 7: There is evidence that average income in the Provider group makes a difference. Not clear if this is the measure being requested for endorsement.

Panel Member 8: ACGs are not free, so the use of a proprietary model makes it harder for people to replicate the measure.

Panel Member 9: Developers used the ACG to adjust for variation in risk profile using age, gender, and diagnosis (clinical risk adjustment). This was applied first before testing income as a social risk factor. After risk adjusting for age, gender, and clinical risk, and limiting by insurance type, they found that income does not significantly impact a patient's total resource use. As a potential practical use case example, the study also evaluated Resource Use provider group performance and found there was no discernible difference in performance when adjusting for income. The analysis considered two different data sources to study income variation, Census tract level data and a commercially licensed data source with more specific income data at the household level. Using Census tract data, a 1% increase in income resulted in a $0.09 increase in total reimbursement, a $0.29 increase in resource use, and $0.18 decrease in price. The results highlight how significantly more the ACG score (clinical risk adjustment) and insurance product impact both the cost and resource use measures. Using the commercially purchased data source, with income by household, a 1% increase in income resulted in an $0.08 increase for total reimbursement, $0.13 increase in resource use, and $0.05 decrease in price. This shows that using a data source that is more specific results in income being less impactful on TCOC and resource use while ACG and product type (commercial or Medicaid) show similar results. They also found a less than 1% change in performance for provider groups when income was added to the RUI model. They considered this insignificant when comparing provider performance. It is not really clear if they included income or not, but I am assuming no given the description of findings related to income.

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Testing attachment, section 2b4.

    Panel Member 1: None.

    Panel Member 2: Yes.

    Panel Member 4: Yes.

    Panel Member 5: No. Practically meaningful difference in performance will vary by use of the measures. This is because some uses may have a higher threshold for differences. For example, a 10% difference in performance when the result is used for public reporting could be very meaningful in terms of provider patient growth and retention strategies. The same 10% difference may not be as meaningful when using the measures internally for improvement work and identification of a work plan.

    Panel Member 6: Submitter included a three page table of raw data - not sure how to assess that. I am assuming we should not do the analysis for them. There is a description of x were better than average, etc. No statistical assessment or summary of meaningful difference analysis.

    Panel Member 7: See previous statement about why NQF is being asked to endorse an internal evaluation/comparison measure for a single health care system.

    Panel Member 8: None.

    Panel Member 9: No concerns.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    Submission document: Testing attachment, section 2b5.

    Panel Member 1: None.

Panel Member 2: N/A

Panel Member 4: The lack of social risk factors is a big omission. However, I'm not clear on the benefit of using social risk adjustors for cost and resource use measures as they have the potential to mask differences. Differences in costs and utilization by social risk factors are important to document and it seems could be handled through stratification.

Panel Member 5: NA

Panel Member 6: NA

Panel Member 7: The Developer did not provide any information in section 2b5 regarding this question.

Panel Member 8: None.

22. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

Panel Member 1: None.

Panel Member 2: Developers claim this is a full population based measure so no missing data. Would like to see some analysis of potentially missing data impact on ACG scores.

Panel Member 4: No concerns.

Panel Member 5: None

Panel Member 6: NA

Panel Member 7: According to the Developer, there are no missing data.

Panel Member 8: None.

For cost/resource use measures ONLY:

23. **Are the specifications in alignment with the stated measure intent?**

☒ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

Panel Member 1: Patients without pharmacy (Rx) benefits are still included in the measures but their Rx cost seems to be imputed by those with Rx benefit. It would have been good to see if the developer conducts a sensitivity analysis in which they estimate the measure with and without the members without the Rx benefit. The truncation of members' cost at $125K is somewhat arbitrary. Why $125K but not $100K?

Panel Member 2: Use assigned when patients have selected/been assigned to provider; attribute using plurality rule. No discussion of proportion attributed or percent care received from attributed provider

Panel Member 4: No concerns

Panel Member 6: All testing performed only on the submitters data - very little discussion of the generalizability of the results.

Panel Member 7: The RUI measure is described as a "population" measure, but is computed at a subpopulation (Provider) level. The description is misleading.

Panel Member 8: This measure captures resource use - method for imputation makes sense.

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Panel Member 1: My validity rating is based on my assessment in #18 through #24 above.

Panel Member 2: Measure well assessed. Some missing discussion around attribution leads to moderate score.

Panel Member 3: The analysis does not address the relationship of the measure to the quality construct

Panel Member 4: The strongest rationale for validity of this measure was correlations for the ACGs, non-risk adjusted PMPMs, and non-risk adjusted TCRRVs with the utilization measures.

Panel Member 5: Extensive assessment of measure components and sound approaches.

Panel Member 6: Concerned about the lack of assessment submitters patient cohort vs that used to calibrate the ACGs. Lack of analysis of the ability to identify meaningful differences.

Panel Member 7: As stated elsewhere, why is NQF endorsing an internal measure used by a single health care system? As a single system, internal measure, the measure seems to function adequately for the use of that system.

Panel Member 8: Strong evidence that measure is capturing average resource use for patients seen by a given provider group.

Panel Member 9: Results above show strong validity of the measure.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Panel Member 2: None. The large minimum sample size of 600 addresses many concerns about reliability seen in measures with lower thresholds.

Panel Member 6: Should measures use proprietary risk adjustment tools (ACG)?

Panel Member 7: Can this measure be adapted so that a true RUI (not costs, but resource use and intensity) could be computed for all major health care systems? If so, then the measure may have some

general value to compare health care resource use across the US. Additionally, the exclusion of the >65 population from the analysis is troubling, given that this is a fast-growing segment of the population and a known high-user group for health care resources.

Developer Submission

Brief Measure Information

NQF #: 1598

De.2. Measure Title: Total Resource Use Population-based PMPM Index

Co.1.1. Measure Steward: HealthPartners

De.3. Brief Description of Measure: 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.

IM.1.1. Developer Rationale: By measuring population based relative resource use, health plans and providers can improve the affordability of health care without sacrificing quality. HealthPartners' RUI gives provider groups valuable information on resource use and, when viewed in conjunction with quality metrics, information on the efficiency of care. The HealthPartners RUI measure is a population-based, patient-centered, total resource use measure, created with Total Care Relative Resource Values that cross all categories of health services. This is in contrast to the many, episodic based resource use measures available in the market today. Both population based and episodic based resource use measures are important and complementary but a key benefit of population based measures is helping to better understand potential overuse & underuse (e.g., although efficient at spine surgery, may be performing too many).

De.1. Measure Type: Cost/Resource Use

S.5. Data Source: Claims

S.3. Level of Analysis: Clinician : Group/Practice, Population : Community, County or City

IF Endorsement Maintenance – Original Endorsement Date: Jan 31, 2012 **Most Recent Endorsement Date:** Jul 13, 2017

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall, less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

IM.1. Opportunity for Improvement

IM.1.1. Briefly explain the rationale for this measure (e.g., the benefits or improvements in performance envisioned by use of this measure)

By measuring population based relative resource use, health plans and providers can improve the affordability of health care without sacrificing quality. HealthPartners' RUI gives provider groups valuable information on resource use and, when viewed in conjunction with quality metrics, information on the efficiency of care. The HealthPartners RUI measure is a population-based, patient-centered, total resource use measure, created with Total Care Relative Resource Values that cross all categories of health services. This is in contrast to the many, episodic based resource use measures available in the market today. Both population based and episodic based resource use measures are important and complementary but a key benefit of population based measures is helping to better understand potential overuse & underuse (e.g., although efficient at spine surgery, may be performing too many).

IM.1.2. Provide performance scores on the measure as specified (current and over time) **at the specified level of analysis.** (This is required for endorsement maintenance. Include mean, stddev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include). **This information also will be used to address the subcriterion on improvement (U.3.1.) under Usability and Use.**

HealthPartners primary care network (Minnesota, western Wisconsin, northeast Wisconsin, Iowa, North Dakota, South Dakota) consists of 65 individual provider groups that have 1,200 clinic sites. Provider group size vary from 600 to a few large systems with 50,000+ members. Please see 2b4.2 of the testing attachment for performance scores across measured provider groups highlighting variation of total resource use (2017 – 2019 dates of service).

Performance is measured on an Index basis relative to 1.00 where each one point (0.01) variation from 1.00 (average) represents a 1% deviation from average. Statistical significance ranges of performance are not necessary as the measure is based on the full population. The results can be analyzed by percentile, percent from mean, standard deviation and clustering methods, this is dependent upon the business application of the measure.

Practically meaningful difference in performance will vary by use of the measures. This is because some uses may have a higher threshold for differences. For example, a 5% increase in a providers' year over year performance would indicate a significant change, whereas a provider performing 5% higher than average might be considered average when developing a tiered network benefit.

The following will give a general sense of the dispersion of the scoring:

Out of the 65 provider groups measured in Total Resource Use measure:

- 26 were better than average
- 3 were 10% better than average
- 12 were 10% higher than average
- 50 were within 10% of the average

IM.1.3. If no or limited performance data on the measure as specified is reported in IM.1.2., then provide a summary of data from the literature that indicates opportunity for improvement or overall, less than optimal performance on the specific focus of measurement.

Performance data was provided. Not applicable.

IM.1.4. Provide disparities data from the measure as specified (current and over time) **by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** (This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) **This information also will be used to address the subcriterion on improvement (U.3.1.) under Usability and Use.**

As previously described in this application, the measure is being submitted for a commercially insured population. Therefore, performance by insurance status is not applicable because the population is all commercially insured. The clinical risk adjustment process described in 2b3. describes how age and gender are accounted for in the methodology and no additional measure performance was tested because this is not how they are being used. That said, in looking at single specialty obstetric and pediatric groups, we see a uniformly distributed result across our network performance and these groups are not clustered, which demonstrates results are not biased against age or gender. Additionally, this demonstrates the clinical risk adjustment is working effectively. The measure is used as a population-based method primarily for payment, benefit design, transparency and improvement.

After applying clinical risk adjustment, socioeconomic testing was conducted on a combined commercial and Medicaid population that considered income and education status as potential factors beyond those already adjusted for. Medicaid was included to ensure we captured a wide range of income levels.

Model Results

1% Income Increase:

Total Reimbursement $0.09

Resource Use $0.28

Price $(0.18)

1% ACG Increase:

Total Reimbursement $4.90

Resource Use $5.00

Price $0.05

Commercial vs. Medicaid Membership:

Total Reimbursement $207.41

Resource Use $(55.39)

Price $261.07

Resource Use Endorsed Measure R2 = 0.5761

Resource Use Endorsed Measure + Income R2 = 0.5758

Using Census tract data, a 1% increase in income resulted in a $0.09 increase in total reimbursement, a $0.29 increase in resource use, and $0.18 decrease in price. The results highlight how significantly more the ACG score (clinical risk adjustment) and insurance product impact both the cost and resource use measures. For frame of reference, on average for the Midwest market, the total spend for a member per month (PMPM) is $500. The results of the evaluation show that a 1% increase in risk score accounts for a $4.90 or roughly 1% increase in PMPM.

Product also contributed significantly with there being a $207 dollar difference in cost between commercial and Medicaid. The variation in resource use was much less, however, still significant with Medicaid covered members utilizing $55 more dollars of resources. The fact that Medicaid's cost per service is approximately half that of commercial rates drives the differences between the TCOC and Resource Use results. The R2 results further emphasize that ACG score and insurance type are the main drivers of cost and resource use variation and income does not provide any additional predictive power.

Methodology and testing results can be found here:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-ses-social-risk-factors-technical-documentation-2020-2021.pdf

IM.1.5. If no or limited data on disparities from the measure as specified is reported in IM.1.4., then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Not applicable

IM.2. Measure Intent

IM.2.1. Describe intent of the measure and its components/ Rationale (including any citations) for analyzing variation in resource use in this way.

Key considerations when constructing the measure:

• The purpose of population-based measurement is to better understand overuse, underuse, and person-centered management and accountability

• Population based-measurement nicely complements condition and episode-base measures; combined they depict a complete picture of a provider's resource use

• Risk adjustment is a critical component to the measure to allow for fair comparisons

• Use this measure as part of a Triple-aim approach where Total Resource Use is a complement to total cost of care, quality and patient experience.

• Removing price via Total Care Relative Resource Values (TCRRVs) allows for a clear picture of resource use opportunities.

• The Resource Use Index measure, when used with a Total Cost of Care measure, will help to better understand cost and resource use opportunities.

Scientific Acceptability of Measure Properties

Extent to which the measure, **as specified**, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific *(check all the areas that apply):*

Care Coordination

Safety : Overuse

De.7. Care Setting *(Select all the settings for which the measure is specified and tested):*

Emergency Department and Services

Home Care

Inpatient/Hospital

Other: All care settings included

Outpatient Services

Post-Acute Care

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

<WebPageURLExists node Type="1">There were no measure specification changes since re-endorsement in 2017. Current endorsed measure materials reside here: https://www.healthpartners.com/about/improving-healthcare/tcoc/, Technical Documentation located here: https://www.healthpartners.com

S.2. Type of resource use measure *(Select the most relevant)*

Per capita (population- or patient-based)

S.3. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED):*

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

S.4. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.5. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.5.1.

Claims

S.5.1. Data Source or Collection Instrument *(Identify the specific data source or data collection instrument, e.g., name of database, clinical registry, collection instrument, etc.)*

- Users administrative claims data base
- Risk Adjustment Tool, Johns Hopkins ACG System
- Standardized costing code table, Total Care Relative Resource Values (TCRRV) specification provided

S.5.2. Data Source or Collection Instrument Reference *(available at measure-specific Web page URL identified in S.1 OR in the file attached here) (Save file as: S_5_2_DataSourceReference)*

S.6. Data Dictionary or Code Table *(Please provide a web page URL or attachment if exceeds 2 pages. NQF strongly prefers URLs. Attach documents only if they are not available on a web page.)*

Data Dictionary:

URL:

Please supply the username and password:

Attachment:

Code Table:

URL:
https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/dev_057916.pdf

Please supply the username and password:

Attachment:

Construction Logic

S.7.1. Brief Description of Construction Logic

If applicable, summarize the general approach or methodology to the measure construction. This is most relevant to measures that are part of or rely on the execution of a measure system or applies to multiple measures.

The measure examines total resource use of a commercial population for a given measurement year (e.g., January 1 to

December 31), for all members eligible for the measure.

S.7.2. Construction Logic *(Detail logic steps used to cluster, group or assign claims beyond those associated with the measure's clinical logic.)*

• All claims included in the measure have a date of service in the measurement year (e.g., between January 1 and December 31)

• Members have a minimum 9 months enrollment in the measurement year

• Commercial population only

• Attribution

• Costing Method – Total Care Relative Resource Values TCCRVs

• Risk Adjustment

S.7.2a. CONSTRUCTION LOGIC ATTACHMENT or URL: If needed, attach supplemental documentation (Save file as: S_7_2_Construction_Logic). All fields of the submission form that are supplemented within the attachment must include a summary of important information included in the attachment and its intended purpose, including any references to page numbers, tables, text, etc.

URL: https://www.healthpartners.com/content/dam/brand-identity/pdfs/plan/tcoc-technical-implementation-guidelines.pdf

Please supply the username and password:

Attachment:

S.7.3. Concurrency of clinical events, measure redundancy or overlap, disease interactions *(Detail the method used for identifying concurrent clinical events, how to manage them, and provide the rationale for this methodology.)*

We do not provide specifications for concurrency of clinical events.

Not applicable. This is a population-based measure that applies to all care settings and conditions.

S.7.4. Complementary services *(Detail how complementary services have been linked to the measure and provide rationale for this methodology.)*

We do not provide specifications for linking complementary services.

Not applicable. This is a population-based measure that applies to all care settings and conditions.

S.7.5. Clinical hierarchies *(Detail the hierarchy of codes or condition groups used and provide rationale for this methodology.)*

We do not provide specifications for clinical hierarchies.

S.7.6. Missing Data *(Detail steps associated with missing data and provide rationale for this methodology (e.g., any statistical techniques to impute missing data)*

We do not provide measure specifications or guidelines for missing data :

In the instances where members have pharmacy benefit carve-outs the following methodology is applied.

The Resource Use measure accounts for members that have their pharmacy benefit carved out by using the members that have pharmacy coverage as a proxy. This technique allows for members without pharmacy coverage to be included in the medical portion of the total resource use with their pharmacy TCRRVs reflecting the provider's risk adjusted pharmacy TCRRVs from those covered. The measures separate the total resource use into medical and pharmacy and only includes the members with pharmacy coverage into the TCRRV PMPM calculation for pharmacy. The total TCRRV PMPM for a provider group is then calculated by adding the medical

TCRRV PMPM to the pharmacy TCRRV PMPM: Total TCRRV PMPM = (Medical TCRRVs / Medical MMs) + (Pharmacy TCRRVs / Pharmacy MMs). MM = member months.

HealthPartners' data includes all medical and mental health care. It also includes the majority of pharmacy claims with the exception of some carveouts. The methodology described above was used for testing. If users have additional carve-outs (e.g., mental health) the lowest common denominator principle (i.e. for any given user if their data includes a carve-out for one their method must apply a carve-out for all) needs to be applied to ensure providers are evaluated fairly. This would require all services that are carved out of one segment of input data to be carved out of the measure for all segments of input data and all input components of the measure (e.g. TCRRV PMPMs, attribution, and risk adjustment).

S.7.7. Resource Use Service Categories (Units) (Select all categories that apply)

Inpatient services: Inpatient facility services

Inpatient services: Evaluation and management

Inpatient services: Procedures and surgeries

Inpatient services: Imaging and diagnostic

Inpatient services: Lab services

Inpatient services: Admissions/discharges

Inpatient services: Labor (hours, FTE, etc.)

Other inpatient services

Ambulatory services: Outpatient facility services

Ambulatory services: Emergency Department

Ambulatory services: Pharmacy

Ambulatory services: Evaluation and management

Ambulatory services: Procedures and surgeries

Ambulatory services: Imaging and diagnostic

Ambulatory services: Lab services

Ambulatory services: Labor (hours, FTE, etc.)

Other ambulatory services

Durable Medical Equipment (DME)

Other services not listed

All care is included

All care is included

All care is included

S.7.8. Identification of Resource Use Service Categories (Units)

(For each of the resource use service categories selected above, provide the rationale for their selection and detail the method or algorithms to identify resource units, including codes, logic and definitions.)

The Total Resource Use considers 100% of health care services in the Resource Use Index and is calculated on a risk-adjusted paid per member per month basis as well as benchmarked to a peer group. The Total Care Relative Resource Values (TCRRVTM) is inclusive of both plan and member liability. Detailed identification of units is available in the Total Care Relative Resource Value White Paper.
https://www.healthpartners.com/content/dam/brand-identity/pdfs/plan/tcoc-relative-resource-value.pdf

S.7.8a. If needed, provide supplemental resource use service category specifications in either URL (preferred) or as an attachment (Save file as S.7.8a_RU_Service_Categories):

URL:

Please supply the username and password:

Attachment:

Clinical Logic

S.8.1. Brief Description of Clinical Logic (Briefly describe your clinical logic approach including clinical topic area, whether or not your account for comorbid and interactions, clinical hierarchies, clinical severity levels and concurrency of clinical events.)

Not applicable. This is a population-based measure that applies to all care settings and conditions.

S.8.2. Clinical Logic *(Detail any clustering and the assignment of codes, including the grouping methodology, the assignment algorithm, and relevant codes for these methodologies.)*

Not applicable. This is a population-based measure that applies to all care settings and conditions.

S.8.3. Evidence to Support Clinical Logic Described in S.8.2 *Describe the rationale, citing evidence to support the grouping of clinical conditions in the measurement population(s) and the intent of the measure (as described in IM3)*

Not applicable. This is a population-based measure that applies to all care settings and conditions.

S.8.3a. CLINICAL LOGIC ATTACHMENT or URL: If needed, attach supplemental documentation (Save file as: S_8_3a_Clinical_Logic). All fields of the submission form that are supplemented within the attachment must include a summary of important information included in the attachment and its intended purpose, including any references to page numbers, tables, text, etc.

URL:

Please supply the username and password:

Attachment: Appendix_A_TCOC_1604_and_Resource_Use_1598_2016_2017_Submission_Testing.pdf

S.8.4. Measure Trigger and End mechanisms *(Detail the measure's trigger and end mechanisms and provide rationale for this methodology)*

All claims dates of service in the measurement year (e.g. January 1 – December 31)

S.8.5. Clinical severity levels *(Detail the method used for assigning severity level and provide rationale for this methodology)*

We do not provide specifications for clinical severity levels.

This is accounted for in application of risk adjustment, Johns Hopkins, ACG System

S.8.6. Comorbid and interactions *(Detail the treatment of co-morbidities and disease interactions and provide rationale for this methodology.)*

We do not provide specifications for co-morbidies and disease interactions.

This is accounted for in application of risk adjustment, Johns Hopkins, ACG System

Adjustments for Comparability

S.9.1. Inclusion and Exclusion Criteria *Detail initial inclusion/exclusion criteria and data preparation steps (related to clinical exclusions, claim-line or other data quality, data validation, e.g., truncation or removal of low or high dollar claim, exclusion of ESRD patients)*

:

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

- Members over age 64
- Members under age 1
- Member enrollment less than nine months during the one year measurement time window
- TCRRVs per member up to 125,000 are included; TCRRVs per member above 125,000 are excluded (truncated)
- Administrative claims covering all categories of health care services: professional, facility inpatient and outpatient, pharmacy,

lab, radiology and any other ancillary healthcare services are included.

- Johns Hopkins ACG System for risk adjustment
- Membership eligibility, identifier and number of months during the measurement period the member was eligible (member

months)

The following should be reviewed prior to beginning implementation of the Total Resource Use measure to ensure data comparability:

- Consistent population of primary and secondary claims diagnosis. Population prevalence to ensure reasonable/completeness of

disease; primary and secondary diagnosis are consistently populated (e.g., diagnosis 1 - 4)

- Data elements are populated within reasonable tolerances and thresholds (e.g., expected CPT ranges, expected allowed amount

ranges, expected units ranges)

- All service categories are available and appropriately represented (e.g., inpatient, pharmacy, outpatient and professional)
- Peer group/case-mix need to be comparable
- Risk adjustment weight and application must be in sync (e.g., truncation threshold values)

It is recommended that further reliability and validity testing be conducted if the user varies from the "Technical Guidelines"

provided. Examples include:

- The user implements the measure with less than 600 members attributed to a provider
- The user applies a different unit of evaluation, such as an employer group, condition or community rather than a provider
- The user employs an alternative attribution algorithm or risk adjustment tool

Paid medical and pharmacy administrative claims for the measurement year (e.g., between January 1 and December 31), allowing

for three months of run out for claims lag.

In the instances where members have pharmacy benefit carve-outs the following methodology is applied.

The Resource Use measure accounts for members that have their pharmacy benefit carved out by using the members that have pharmacy coverage as a proxy. This technique allows for members without pharmacy

coverage to be included in the medical portion of the total resource use with their pharmacy TCRRVs reflecting the provider's risk adjusted pharmacy TCRRVs from those covered. The measures separate the total resource use into medical and pharmacy and only includes the members with pharmacy coverage into the TCRRV PMPM calculation for pharmacy. The total TCRRV PMPM for a provider group is then calculated by adding the medical TCRRV PMPM to the pharmacy TCRRV PMPM: Total TCRRV PMPM = (Medical TCRRVs / Medical MMs) + (Pharmacy TCRRVs / Pharmacy MMs). MM = member months.

HealthPartners' data includes all medical and mental health care. It also includes the majority of pharmacy claims with the exception of some carveouts. The methodology described above was used for testing. If users have additional carve-outs (e.g., mental health) the lowest common denominator principle (i.e., for any given user if their data includes a carve-out for one their method must apply a carve-out for all) needs to be applied to ensure providers are evaluated fairly. This would require all services that are carved out of one segment of input data to be carved out of the measure for all segments of input data and all input components of the measure (e.g., TCRRV PMPMs, attribution, and risk adjustment).

S.9.2. Risk Adjustment Type (Select type)

Statistical risk model

If other:

S.9.3. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets)*

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

S.9.4 Costing method

Detail the costing method including the source of cost information, steps to capture, apply or estimate cost information, and provide rationale for this methodology.

Standardized pricing

Standardized pricing

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. 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.

General Overview of Application:

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.

Detail development:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/plan/tcoc-relative-resource-value.pdf

Sample TCRRV table:

https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/dev_057916.pdf

S.10. Type of score*(Select the most relevant):*

Ratio

Other (specify):

If other: https://www.healthpartners.com/content/dam/brand-identity/pdfs/plan/tcoc-sample-medical-group.pdf page 9

Attachment:

S.11. Interpretation of Score *(Classifies interpretation of a ratio score(s) according to whether higher or lower resource use amounts is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score, etc.)*

A provider Total Resource Use Index (RUI) of 1.10 equates to 10% higher risk adjusted resource use. Similarly, a provider RUI score of 0.90 equates to 10% less paid risk adjusted resource use.

A score of 1.0 is equivalent to the peer group average.

S.12. Detail Score Estimation *(Detail steps to estimate measure score.)*

There is no estimation in the Total Resource Use Measure. The actual result is calculated as follows:

Resource Use Index (RUI):

Numerator: Total Resource PMPM = (Total Medical TCRRV / Medical Member Months) + (Total Pharmacy TCRRV / Pharmacy Member Months)

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

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

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

Reporting Guidelines

This section is optional and will be available for users of the measure as guidance for implementation and reporting.

S.13.1. Describe discriminating results approach

Detail methods for discriminating differences (reporting with descriptive statistics--e.g., distribution, confidence intervals).

This is a full population-based measure, therefore, confidence intervals are not applicable. The results can be analyzed by percentile, percent from mean, standard deviation and clustering methods, this is dependent upon the business application of the

measure.

A provider Total Resource Use Index (RUI) of 1.10 equates to 10% higher risk adjusted resource use. Similarly, a provider RUI score of 0.90 equates to 10% less paid risk adjusted resource use.

A score of 1.00 is equivalent to the peer group average.

S.13.2. Detail attribution approach

Detail the attribution rules used for attributing resources/costs to providers (e.g., a proportion of total measure cost or frequency of visits during the measure's measurement period) and provide rationale for this methodology.

There are three main options to include members in the Total Resource Use measure, by geographic region or defined population, assignment of members to a responsible party or attribution of members to a responsible party. Each option will require a different approach for assigning members to the responsible party or unit of measurement. In all cases the measure exclusions will still need to be applied (i.e., age 1-64, commercial members, enrolled a minimum of 9 months).

- Population – categorize members by geographic areas such as state, county or zip code or in a defined population, for example an employer group.
- Assignment – each member is assigned to a responsible provider group regardless of where he or she received care
- Attribution – in instances where members are not assigned a responsible provider and are able to receive care at any provider (an open access market), applying an Attribution Algorithm will be necessary to identify the provider that is responsible for managing the patient's care. There are several options for defining the responsible provider, but in general the provider that sees the patient the most often is attributed the member

For purposes of testing, a commonly used Attribution Algorithm for an open access market was applied - plurality model, using most recent visit as a tie breaker:

- Include all professional claims experience claims data in a twelve month measurement period (e.g., January 1 – December 31)
- Exclude all services that are not office based
- Exclude convenience care clinic visits
- Exclude all providers that are not a physician, physician assistant or nurse practitioner
- Assign each service line a specialty based on the servicing physician's practicing specialty or credential specialty if practicing specialty is not available
- Include only the following specialties: Family Medicine, Internal Medicine, Pediatrics, Geriatrics, OB/GYN

A real-case example of using our Total Resource Use measure at a population level:

- The Network for Regional Healthcare Improvement (NRHI), a licensee of HealthPartners' measures, released their first report in their Getting to Affordability initiative in November 2016 called "From Claims to Clarity: Deriving Actionable Healthcare Cost Benchmarks from Aggregated Commercial Claims Data". This report compared commercially insured health plan members across five geographic regions. When looking at the population level (regional geographic view), all members with commercial insurance were included, yet when the individual regions were evaluated at the provider group practice level, they used attribution to only include patients who had a visit with a primary care provider.

Technical specifications:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/plan/tcoc-technical-implementation-guidelines.pdf

HealthPartners has studied various attribution methods, our findings are located here: HealthPartners Attribution Technical Paper

https://www.healthpartners.com/content/dam/brand-identity/pdfs/plan/healthpartners-attribution-technical-paper.pdf

S.13.3. Identify and define peer group

Identify the peer group and detail how peer group is identified and provide rationale for this methodology.

The peer group can be applied by market, region or national with the following criteria:

- Provider Specialties include: Internal Medicine, Family Medicine, Pediatrics, Geriatrics and OB/GYN
- Provider Types include: Physician, Physician Assistant, Nurse Practitioner

S.13.4. Sample size

Detail the sample size requirements for reporting measure results.

This measure has been tested for a minimum attributed member population of 600 members, this number is aligned with community-based quality and patient experience measures in the market tested. We recommend further reliability and validity testing if a threshold less than 600 attributed members is used.

S.13.5. Define benchmarking and comparative estimates

Detail steps to produce benchmarking and comparative estimates and provide rationale for this methodology.

The Resource use measure is relative to a benchmark or peer group of the user's choice. This can be a group of members, providers, geographic regions or any grouping of member data. The idea is that 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 peer group average is set as the benchmark and a provider's Total Resource Use ACG Adjusted PMPM is indexed against the peer group average. The Peer Group average is calculated in the same manner as an individual provider:

Resource Use (RUI):

Numerator: Peer Group Total Resource Use PMPM = (Peer Group Total Medical TCRRV/ Peer Group Medical Member Months) + (Peer Group Total Pharmacy TCRRV / Peer Group Pharmacy Member Months)

Denominator: Peer Group ACG Risk Score

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

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

Validity – See attached Measure Testing Submission Form

SA.1. Attach measure testing form

1598_Total_Resource_Use_v7.1_nqf_testing_attachment_4.22.21.docx

Measure Number (*if previously endorsed*)**:** 1604
Measure Title: Total Cost of Care Population based PMPM Index
Date of Submission: 4/16/2021

Type of Measure:

Measures	Measures (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☒ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. ***If there are differences by aspect of testing****,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.*

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for **all** the sources of data specified and intended for measure implementation. **If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Commercial administrative claims

Medicaid administrative claims were used in addition to commercial claims for purposes of socioeconomic status (SES) testing to support social risk factor evaluation for this measure.

1.3. What are the dates of the data used in testing?
2018, 2019 dates of service for validity testing

2019 dates of service for reliability testing

2019 dates of service for SES/social risk factor testing

1.4. What levels of analysis were tested? (*testing must be provided for **all** the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

HealthPartners primary care network (Minnesota, western Wisconsin, northeast Wisconsin, Iowa) consists of 65 individual provider groups that have 1,200 clinic sites. Provider group size vary from 600 to a few large systems with 50,000+ members.

Prior submission: HealthPartners' primary care network providers (2014 and 2015).

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

This is a population-based measure that applies to all care settings and conditions using HealthPartners health plan's full book of business. The total membership of the primary care attributed network is over 555,000 members in 2019.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Reliability and Validity testing use the same population and underlying data. The social risk factor analysis also includes the Medicaid population.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g., census tract), or patient community characteristics (e.g., percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

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 to commercial only. Social risk factors analyzed, specifically socioeconomic testing was conducted that considered income and education status as potential factors beyond those already adjusted for.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

Prior submission: Please see Appendix A (page 2) for reliability testing results from prior submission. The method of testing (bootstrapping and 90% random sample) used for current resubmission is the same methodology used in prior submission.

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Overview of Analysis

Resource Use Index (RUI) is a measure of a provider's effectiveness of managing their primary care attributed population across the care continuum. The RUI measure was applied to HealthPartners primary care providers as per the measure specifications and results were calculated for 2019.

The reliability testing demonstrates the repeatability of producing the same results a high proportion of the time. To measure the reliability of the RUI, measure the actual results were compared to the results calculated by two sampling methods, bootstrapping and a 90% random sample.

These methods were chosen as they represent the measure intent, which is that the RUI measure represents providers' average resource use across their population. Since the measure is aggregated to the provider group level, evaluation of member level variability is not necessary.

In the bootstrapping method members that were attributed to a provider group were randomly selected with replacement. This method artificially creates variation around a provider group's total resource use as each randomly selected iteration (sample populations) does not truly represent the provider's case mix of patients. What this method does however is give an indication as to the repeatability of the measure by comparing how closely the actual resource use measure is to the bootstrapped averages.

In the 90% random sample method, the members that were attributed to a provider group were randomly sampled at the 90% membership level without replacement. This technique was employed to create variation within a provider group by leveraging their own population and controlling for the patient case mix variation that is introduced when random sampling is employed.

Methodology

To perform the bootstrap, the SAS PROC SURVEYSELECT procedure with the Unrestricted Random Sample option for full replacement was utilized to create a series of random samples for each provider group being measured. Full replacement means that one observation is drawn at random, recorded, and then placed back into the data pool so that it can be drawn again if randomly selected. The numbers of records sampled are drawn such that the samples created are the same size as the original number of attributed members for the provider group. In this way, it is theoretically possible (although virtually improbable) to produce a sample of size n that could consist of the same record drawn n times in a row. This sample process was performed 500 times for each provider group being analyzed, to produce 500 sets of risk-adjusted Total Resource Use results for each provider in the analysis.

Once the 500 samples were created for each provider group, the total resource use of each sample for each provider group was compared to the network average to produce a risk adjusted index. The mean Total

Resource Use (RUI) from these 500 iterations was computed and compared to the Actual Resource Use Index (RUI) index for each provider group.

In the second method, 90% of attributed provider group members were randomly selected, without replacement. A 90% sample was used despite having the full health plan provider population, as a concession to provider claims that errors in administrative data may not allow for a perfect 100% representation of their population. The sampling process was performed using the SAS PROC SURVEYSELECT procedure with the Simple Random Sample (SRS) option. This method allows for each attributed member to be selected only one time until 90% of the total provider population has been reached. The 90% sampling process was repeated 500 times for each provider group analyzed. Attributed members' total resource use was aggregated in each sample to produce 500 Total Resource Use Index results for each provider group. The mean of the sampled Total Resource Use index was calculated for each provider group and compared to the Actual RUI index for each provider group.

The bootstrap results should indicate that the within provider RUI variation is significantly less than the between provider variation.

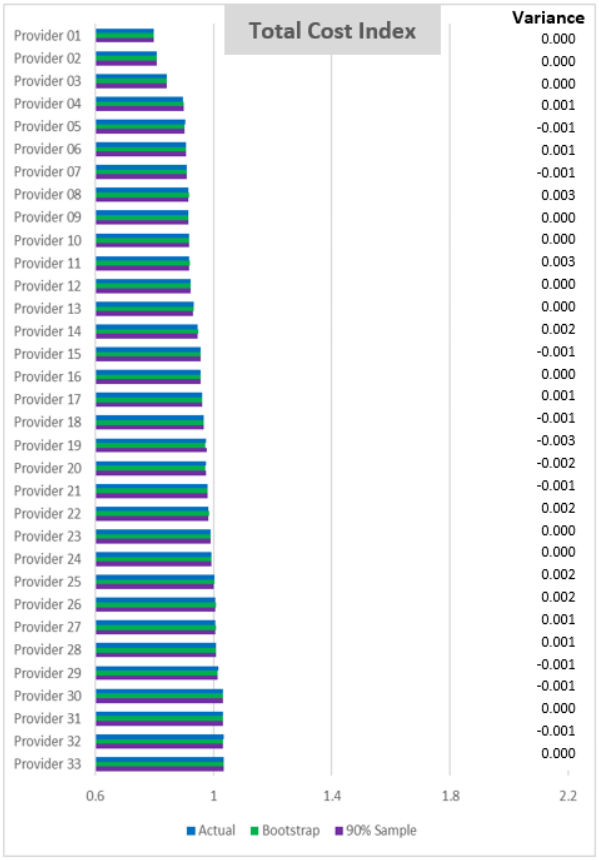
Reliability Paper includes the same method of testing described above:
https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-bootstrap-reliability-testing-2020-2021-cell-based.pdf

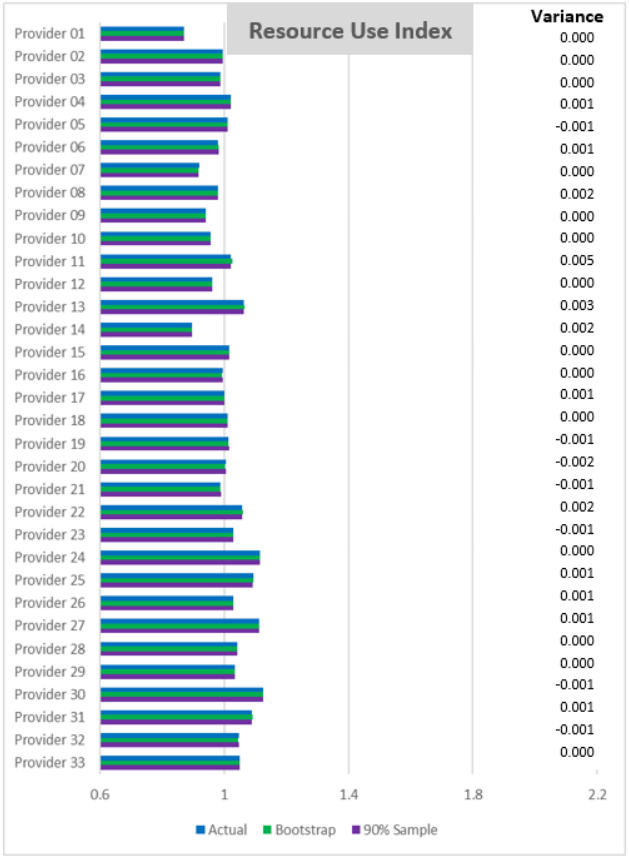
2a2.3. For each level of testing checked above, what were the statistical results from reliability testing?
(e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

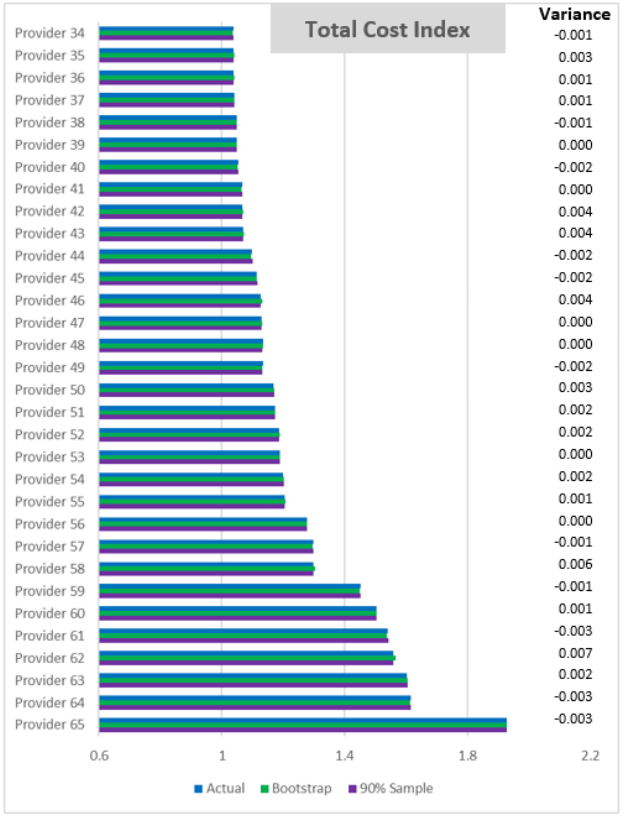
The variances from Actual RUI ranged from -0.0037 to 0.0062 in the bootstrap to -0.0019 to 0.0016 in the 90% sample. The mean Total Resource Use results from the bootstrap and 90% samples compared to the actual RUI results for each provider group are displayed on the charts on the following pages. The variance between the actual RUI to the bootstrap results is shown on the far right of each chart. The RUI charts are sorted in ascending order by Total Cost Index. See Reliability Paper for detailed results.

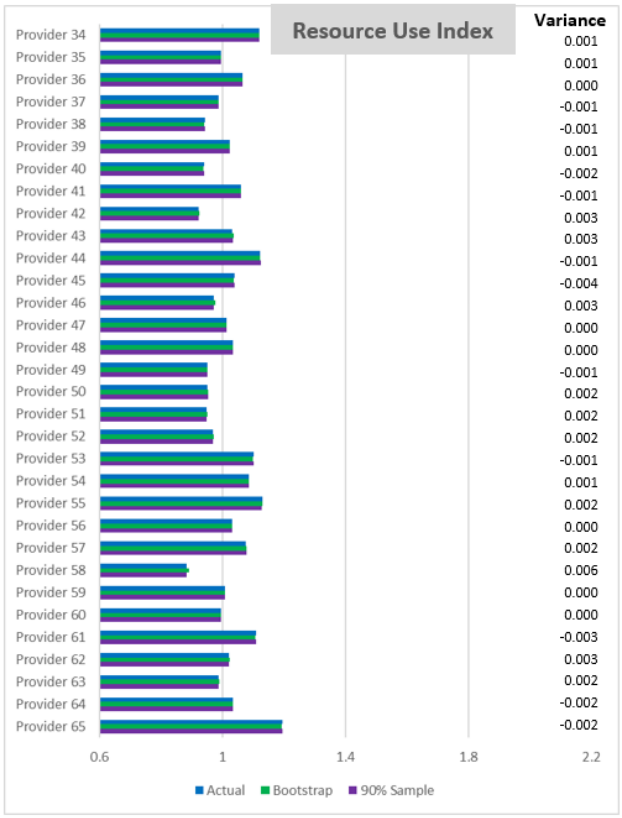
Reliability Paper describes the results of testing in detail:
https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-bootstrap-reliability-testing-2020-2021-cell-based.pdf

Prior submission: Please see Appendix A (page 2) for prior submission results.


Total Cost Index
Variance
Provider 01 0.000
Provider 02 0.000
Provider 03 0.000
Provider 04 0.001
Provider 05 -0.001
Provider 06 0.001
Provider 07 -0.001
Provider 08 0.003
Provider 09 0.000
Provider 10 0.000
Provider 11 0.003
Provider 12 0.000
Provider 13 0.000
Provider 14 0.002
Provider 15 -0.001
Provider 16 0.000
Provider 17 0.001
Provider 18 -0.001
Provider 19 -0.003
Provider 20 -0.002
Provider 21 -0.001
Provider 22 0.002
Provider 23 0.000
Provider 24 0.000
Provider 25 0.002
Provider 26 0.002
Provider 27 0.001
Provider 28 0.001
Provider 29 -0.001
Provider 30 -0.001
Provider 31 0.000
Provider 32 -0.001
Provider 33 0.000
0.6
1
1.4
1.8
2.2
Actual
Bootstrap
90% Sample


Resource Use Index
Variance
Provider 01 0.000
Provider 02 0.000
Provider 03 0.000
Provider 04 0.001
Provider 05 -0.001
Provider 06 0.001
Provider 07 0.000
Provider 08 0.002
Provider 09 0.000
Provider 10 0.000
Provider 11 0.005
Provider 12 0.000
Provider 13 0.003
Provider 14 0.002
Provider 15 0.000
Provider 16 0.000
Provider 17 0.001
Provider 18 0.000
Provider 19 -0.001
Provider 20 -0.002
Provider 21 -0.001
Provider 22 0.002
Provider 23 -0.001
Provider 24 0.000
Provider 25 0.001
Provider 26 0.001
Provider 27 0.001
Provider 28 0.000
Provider 29 0.000
Provider 30 -0.001
Provider 31 0.001
Provider 32 -0.001
Provider 33 0.000
0.6
1
1.4
1.8
2.2
Actual
Bootstrap
90% Sample


Total Cost Index
Variance
Provider 34 -0.001
Provider 35 0.003
Provider 36 0.001
Provider 37 0.001
Provider 38 -0.001
Provider 39 0.000
Provider 40 -0.002
Provider 41 0.000
Provider 42 0.004
Provider 43 0.004
Provider 44 -0.002
Provider 45 -0.002
Provider 46 0.004
Provider 47 0.000
Provider 48 0.000
Provider 49 -0.002
Provider 50 0.003
Provider 51 0.002
Provider 52 0.002
Provider 53 0.000
Provider 54 0.002
Provider 55 0.001
Provider 56 0.000
Provider 57 -0.001
Provider 58 0.006
Provider 59 -0.001
Provider 60 0.001
Provider 61 -0.003
Provider 62 0.007
Provider 63 0.002
Provider 64 -0.003
Provider 65 -0.003
0.6
1
1.4
1.8
2.2
Actual
Bootstrap
90% Sample


Resource Use Index
Variance
Provider 34 0.001
Provider 35 0.001
Provider 36 0.000
Provider 37 -0.001
Provider 38 -0.001
Provider 39 0.001
Provider 40 -0.002
Provider 41 -0.001
Provider 42 0.003
Provider 43 0.003
Provider 44 -0.001
Provider 45 -0.004
Provider 46 0.003
Provider 47 0.000
Provider 48 0.000
Provider 49 -0.001
Provider 50 0.002
Provider 51 0.002
Provider 52 0.002
Provider 53 -0.001
Provider 54 0.001
Provider 55 0.002
Provider 56 0.000
Provider 57 0.002
Provider 58 0.006
Provider 59 0.000
Provider 60 0.000
Provider 61 -0.003
Provider 62 0.003
Provider 63 0.002
Provider 64 -0.002
Provider 65 -0.002
0.6
1
1.4
1.8
2.2
Actual
Bootstrap
90% Sample

Provider Group	Total Cost Index: 90% Sample	Total Cost Index: Bootstrap	Total Cost Index: Actual	Total Cost Index: Variation Between Actual and 90%	Total Cost Index: Variation Between Actual and Bootstrap	Total Resource Use: 90% Sample	Total Resource Use: Bootstrap	Total Resource Use: Actual	Total Resource Use: Variation Between Actual and 90%	Total Resource Use: Variation Between Actual and Bootstrap
Provider 01	0.798	0.799	0.799	(0.000)	(0.000)	0.872	0.872	0.872	(0.000)	(0.000)
Provider 02	0.809	0.809	0.809	0.000	0.000	0.995	0.995	0.996	(0.000)	0.000
Provider 03	0.843	0.844	0.843	(0.000)	(0.000)	0.987	0.988	0.988	(0.000)	(0.000)
Provider 04	0.900	0.899	0.899	0.002	(0.001)	1.023	1.022	1.021	0.002	(0.001)
Provider 05	0.902	0.903	0.904	(0.003)	0.001	1.010	1.011	1.012	(0.002)	0.001
Provider 06	0.907	0.908	0.907	(0.000)	(0.001)	0.982	0.983	0.982	0.000	(0.001)
Provider 07	0.910	0.910	0.911	(0.000)	0.001	0.919	0.919	0.920	(0.001)	0.000
Provider 08	0.917	0.919	0.916	0.001	(0.003)	0.980	0.981	0.979	0.001	(0.002)
Provider 09	0.917	0.917	0.917	0.000	0.000	0.943	0.942	0.942	0.000	0.000
Provider 10	0.917	0.917	0.917	(0.000)	0.000	0.956	0.956	0.956	(0.000)	0.000
Provider 11	0.917	0.921	0.918	(0.000)	(0.003)	1.021	1.027	1.022	(0.001)	(0.005)
Provider 12	0.923	0.923	0.923	(0.000)	(0.000)	0.961	0.961	0.961	(0.000)	0.000
Provider 13	0.932	0.934	0.933	(0.001)	(0.000)	1.063	1.067	1.064	(0.001)	(0.003)
Provider 14	0.947	0.948	0.946	0.000	(0.002)	0.896	0.898	0.896	0.000	(0.002)
Provider 15	0.958	0.957	0.957	0.000	0.001	1.018	1.017	1.018	0.000	0.000
Provider 16	0.958	0.957	0.958	(0.000)	0.000	0.995	0.994	0.995	(0.000)	0.000
Provider 17	0.961	0.964	0.962	(0.001)	(0.001)	1.001	1.002	1.001	(0.000)	(0.001)
Provider 18	0.969	0.968	0.969	0.000	0.001	1.011	1.011	1.011	0.000	0.000
Provider 19	0.977	0.972	0.975	0.002	0.003	1.017	1.014	1.015	0.002	0.001
Provider 20	0.976	0.974	0.975	0.001	0.002	1.007	1.004	1.006	0.001	0.002
Provider 21	0.981	0.979	0.980	0.000	0.001	0.990	0.988	0.989	0.001	0.001
Provider 22	0.984	0.985	0.984	0.000	(0.002)	1.058	1.060	1.058	0.000	(0.002)
Provider 23	0.992	0.992	0.992	0.000	(0.000)	1.031	1.029	1.030	0.000	0.001
Provider 24	0.993	0.992	0.992	0.001	(0.000)	1.116	1.115	1.115	0.001	(0.000)
Provider 25	1.002	1.004	1.003	(0.001)	(0.002)	1.092	1.094	1.094	(0.002)	(0.001)
Provider 26	1.007	1.009	1.006	0.001	(0.002)	1.030	1.030	1.029	0.001	(0.001)

Provider Group	Total Cost Index: 90% Sample	Total Cost Index: Bootstrap	Total Cost Index: Actual	Total Cost Index: Variation Between Actual and 90%	Total Cost Index: Variation Between Actual and Bootstrap	Total Resource Use: 90% Sample	Total Resource Use: Bootstrap	Total Resource Use: Actual	Total Resource Use: Variation Between Actual and 90%	Total Resource Use: Variation Between Actual and Bootstrap
Provider 27	1.007	1.008	1.007	(0.000)	(0.001)	1.112	1.113	1.112	(0.000)	(0.001)
Provider 28	1.009	1.010	1.009	(0.000)	(0.001)	1.042	1.042	1.042	(0.000)	(0.000)
Provider 29	1.015	1.015	1.016	(0.001)	0.001	1.034	1.034	1.035	(0.001)	0.000
Provider 30	1.033	1.033	1.033	(0.000)	0.001	1.125	1.125	1.125	(0.000)	0.001
Provider 31	1.033	1.033	1.033	(0.000)	0.000	1.090	1.091	1.091	(0.000)	(0.001)
Provider 32	1.033	1.033	1.034	(0.001)	0.001	1.047	1.046	1.047	(0.001)	0.001
Provider 33	1.034	1.034	1.034	(0.000)	(0.000)	1.051	1.051	1.051	0.000	(0.000)
Provider 34	1.037	1.037	1.037	0.000	0.001	1.120	1.121	1.120	0.000	(0.001)
Provider 35	1.039	1.041	1.038	0.001	(0.003)	0.996	0.997	0.995	0.000	(0.001)
Provider 36	1.039	1.040	1.039	0.000	(0.001)	1.066	1.066	1.066	0.000	(0.000)
Provider 37	1.041	1.042	1.041	(0.000)	(0.001)	0.988	0.987	0.988	0.000	0.001
Provider 38	1.049	1.047	1.048	0.001	0.001	0.944	0.942	0.943	0.001	0.001
Provider 39	1.049	1.050	1.049	0.000	(0.000)	1.024	1.024	1.023	0.000	(0.001)
Provider 40	1.053	1.051	1.053	0.000	0.002	0.940	0.939	0.941	(0.000)	0.002
Provider 41	1.066	1.065	1.066	0.000	0.000	1.060	1.060	1.060	(0.000)	0.001
Provider 42	1.066	1.069	1.066	(0.000)	(0.004)	0.922	0.925	0.922	(0.000)	(0.003)
Provider 43	1.069	1.073	1.069	0.000	(0.004)	1.033	1.036	1.033	0.000	(0.003)
Provider 44	1.100	1.096	1.097	0.003	0.002	1.126	1.123	1.123	0.002	0.001
Provider 45	1.115	1.112	1.115	0.001	0.002	1.039	1.036	1.039	(0.000)	0.004
Provider 46	1.127	1.131	1.128	(0.000)	(0.004)	0.973	0.976	0.973	(0.000)	(0.003)
Provider 47	1.130	1.131	1.130	(0.000)	(0.000)	1.015	1.015	1.015	(0.000)	0.000
Provider 48	1.133	1.133	1.133	(0.000)	0.000	1.034	1.035	1.034	0.000	(0.000)
Provider 49	1.133	1.132	1.134	(0.001)	0.002	0.952	0.951	0.952	(0.000)	0.001
Provider 50	1.170	1.171	1.169	0.001	(0.003)	0.952	0.953	0.951	0.001	(0.002)
Provider 51	1.173	1.175	1.173	0.000	(0.002)	0.949	0.951	0.949	0.000	(0.002)
Provider 52	1.187	1.188	1.187	0.000	(0.002)	0.970	0.972	0.970	0.000	(0.002)
Provider 53	1.188	1.188	1.188	(0.000)	(0.000)	1.101	1.100	1.101	(0.000)	0.001

Provider Group	Total Cost Index: 90% Sample	Total Cost Index: Bootstrap	Total Cost Index: Actual	Total Cost Index: Variation Between Actual and 90%	Total Cost Index: Variation Between Actual and Bootstrap	Total Resource Use: 90% Sample	Total Resource Use: Bootstrap	Total Resource Use: Actual	Total Resource Use: Variation Between Actual and 90%	Total Resource Use: Variation Between Actual and Bootstrap
Provider 54	1.201	1.202	1.200	0.001	(0.002)	1.086	1.087	1.086	0.000	(0.001)
Provider 55	1.204	1.206	1.205	(0.001)	(0.001)	1.128	1.131	1.129	(0.002)	(0.002)
Provider 56	1.277	1.277	1.277	0.000	(0.000)	1.032	1.032	1.032	0.000	(0.000)
Provider 57	1.298	1.296	1.297	0.001	0.001	1.077	1.078	1.076	0.001	(0.002)
Provider 58	1.298	1.303	1.298	0.001	(0.006)	0.884	0.891	0.885	(0.000)	(0.006)
Provider 59	1.451	1.450	1.450	0.000	0.001	1.008	1.008	1.008	(0.000)	(0.000)
Provider 60	1.503	1.504	1.503	0.000	(0.001)	0.995	0.995	0.995	0.000	(0.000)
Provider 61	1.541	1.537	1.540	0.001	0.003	1.110	1.107	1.110	0.000	0.003
Provider 62	1.559	1.565	1.558	0.001	(0.007)	1.022	1.025	1.022	0.000	(0.003)
Provider 63	1.605	1.605	1.603	0.002	(0.002)	0.989	0.990	0.988	0.001	(0.002)
Provider 64	1.615	1.612	1.615	(0.000)	0.003	1.035	1.033	1.035	(0.000)	0.002
Provider 65	1.941	1.936	1.939	0.002	0.003	1.197	1.194	1.196	0.001	0.002

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The results of the Bootstrap and Random Sample tests allow us to confidently conclude that the measures will reliably decipher RUI performance between levels of analysis (e.g., provider group). The bootstrap results indicate that the RUIs are reliable as the provider variation within all groups is <1% whereas the variation between groups spans >110%.

Reliability Paper describes the provider group results of testing in detail:
https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-bootstrap-reliability-testing-2020-2021-cell-based.pdf

---

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

Prior submission: Please see Appendix A (page 8) for validity testing results from prior submission. The method of testing used for current resubmission is the same methodology used in prior submission.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

A Validity Analysis was performed on the HealthPartners' Total Resource Use measure which indicates the results accurately reflect the performance levels of provider groups. When used in conjunction with the Total Cost of Care measure, the measure also accurately identifies the price (per unit cost) performance levels of providers.

Detailed testing can be found in the Validity paper: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-validity-analysis-2021-2021.pdf

Critical data elements

Non-risk adjusted correlations between ACG and Total Resource Use, Total Cost Relative Resource Values (resource use) and utilization metrics were calculated.

Performance Measure Score

Risk adjusted Resource Use Index correlations to known risk adjusted utilization metrics were calculated.

Empirical testing of validity and overview of face validity policy and procedure.

An assessment of high and low performing provider groups supports the relationship between risk adjusted utilization metrics and Resource Use Index.

The face validity process is conducted by transparently sharing results and methods with provider groups measured and allowing a 45-day comment period prior to public display of provider group results.

HealthPartners has a Policy and Procedure Review Process and executes it annually with each release of provider groups' performance and measurement results. Disclosure to providers includes:

1. Transparent reporting of measurement methodology

2. Providing comparative performance results with information on statistical reliability to providers
3. Providing an explanation of the results at least 45 days prior to their use in public reporting or business applications
4. Notifying providers of how the information will be used
5. A process by which providers can notify the plan of additional information or corrections

Public reporting of provider group measurement results:

https://www.healthpartners.com/public/cost-and-quality/index.html

Publicly available methods of rate calculations for transparency:

https://www.healthpartners.com/hp/ratings-methods/

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The correlation coefficients are included below for testing validity of the measure components and validity of the Total Resource Use measure. Interpretation accompanies the tables of results below to provide context. However, please reference the paper to follow the complete analytical pathway with context and reasoning to conclude the measure is valid.

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-validity-analysis-2021-2021.pdf

Validity of Measure Components

Correlations Between ACG Score, Non-Risk Adjusted Per Member Per Month (PMPMs), Non-Risk Adjusted Total Cost Relative Resource Values (TCRRVs), and Risk Adjusted RUI

Metric	Correlation Coefficient: ACG	Correlation Coefficient: Non-Risk Adj PMPMs
Non-Risk Adj PMPM	0.67	1.00
Non-Risk Adj TCRRVs	0.95	0.74
ACG Risk Adj TCI	0.00	0.72
ACG Risk Adj RUI	0.11	0.26
Price	-0.06	0.60

There is a high correlation between ACG score and the non-risk adjusted PMPM and TCRRVs which indicates that the non-risk adjusted PMPM and the non-risk adjusted TCRRVs are a good measure of resource use.

There is a low correlation between ACG score and the risk adjusted RUI. This indicates that the risk score of a provider has no impact on a provider's ability to be a high performer.

There is a low correlation between price and ACG because ACGs measure expected resource use whereas price is not affected by the number or intensity of services received.

Correlations Between the Non-Risk Adjusted Place of Service Metrics and Non-Risk Adjusted PMPMs & Non-Risk Adjusted TCRRVs

Non-Risk Adjusted: Service Category Metric	Correlation Coefficient: Non-Risk Adj Service Category PMPMs	Correlation Coefficient: Non-Risk Adj Service Category TCRRVs
Inpatient: Admits/1000	0.75	0.79
Outpatient: ER/1000	0.62	0.55
Outpatient: OP Surgery/1000	0.63	0.83
Professional: E&M/1000	0.56	0.75
Professional: Lab/Path/1000	0.68	0.84
Professional: Std Rad/1000	0.53	0.77
Professional: High Tech Rad/1000	0.70	0.56
Pharmacy: Rx/1000	0.79	0.84

Inpatient: There should be and are strong correlations between the admit rate to the non-risk adjusted PMPMs and non-risk adjusted TCRRVs as the only two factors not measured by the admits are the intensity and unit cost of the services performed.

Outpatient: There should be and are moderate correlations between the ER and outpatient surgery rates to the non-risk adjusted PMPMs and non-risk adjusted TCRRVs as these two utilization metrics combine to encompass approximately 60% of the total outpatient spend.

Professional: There should be and are moderate correlations between the E&M visits, Lab/Path services, standard radiology, and high tech radiology to the non-risk adjusted PMPMs and non-risk adjusted TCRRVs as they represent 47% of the professional spend, but are also good indicators of patients that consume medical services.

Pharmacy: There should be and are strong correlations between the pharmacy prescribing rates to the non-risk adjusted PMPMs and non-risk adjusted TCRRVs as the only factor that is not accounted for in the Rx

prescribing rate metric is the intensity of the drug prescribed. The intensity includes generic usage as well as the variation in cost between drugs.

Since the ACG score, non-risk adjusted PMPMs and non-risk adjusted TCRRVs are a measure of the consumption of health care services, there should be strong correlation between these values and known utilization metrics.

Composite Utilization: A utilization metric was created by weighting each of the underlying utilization metrics within and across places of service by the percent of resources it represents of the total resources by each provider group.

Composite Utilization Metric by Provider Group =

Inpatient (Admit Rate x Inpatient Resource Use %) +
Outpatient (ER rate, OP Surg Rate, High Tech Rad Rate) x Outpatient Resource Use %) +
Professional (E&M rate, Lab/Path Rate, Std Rad) x Professional Resource Use %) +
Pharmacy (Rx rate x Pharmacy Resource Use %)

Non-Risk Adjusted: Metric	Correlation Coefficient: ACG	Correlation Coefficient: Non-Risk Adj PMPMs	Correlation Coefficient: Non-Risk Adj TCRRVs
Composite Utilization	0.91	0.77	0.95

The non-risk adjusted resource composite is highly correlated with ACGs, non-risk adjusted PMPMs and non-risk adjusted TCRRVs.

Validity of Total Resource Use Measure

Correlations Between the Risk Adjusted Place of Service Metrics and TCI, Price, and RUI

Risk Adjusted: Metric	Correlation Coefficient: TCI	Correlation Coefficient: RUI	Correlation Coefficient: Price
Hospital TCI	0.78	NA	NA
Prof TCI	0.71	NA	NA
Rx TCI	0.21	NA	NA
Hospital RUI	NA	0.28	NA
Prof RUI	NA	0.52	NA
Total RUI	0.19		
Hospital Price	NA	NA	0.88
Prof Price	NA	NA	0.91
Total Price	0.93		

- Total Resource Use is correlated with TCI as expected.

- Professional RUI is correlated with overall RUI, supporting the notion primary care providers are integral in the management of total costs and resources.
- Hospital-based RUI has a lower correlation than professional as a lower proportion of patients require hospital-based care.

Correlations Between Risk Adjusted Place of Service Utilization Metrics and Corresponding RUI

Risk Adjusted: Service Category Metric	Correlation Coefficient: Service Category TCIs	Correlation Coefficient: Service Category RUIs
Inpatient: Admit Rate	0.56	0.73
Outpatient: ER Cnt	0.56	0.42
Outpatient: OP Surgery	0.48	0.66
Professional: E&M Visits	0.42	0.63
Professional: Lab/Path	0.37	0.45
Professional: Std Rad	0.42	0.21
Professional: High Tech Rad	0.44	0.32
Pharmacy: Rx Count	0.22	No data
Composite	0.64	0.47

Inpatient: There is a high correlation between the risk adjusted admit rate and the inpatient RUI. This would indicate that the higher the risk adjusted admit rate the more likely a provider will have a higher than average RUI.

Outpatient: There is a moderate correlation between the risk adjusted ER count and the outpatient RUI. This would indicate that the higher the risk adjusted ER counts the more likely a provider will have a higher than average outpatient RUI.

Professional: The professional utilization metrics are moderately correlated to the professional RUI.

This result is as expected because the professional place of service includes a significant amount of services beyond these three utilization measures (other professional services = 53%).

It is also as expected because having higher than average utilization on diagnostic or management-based services does not necessarily indicate a higher resource consuming patient.

Risk Adjusted: Service Category Metric	Correlation Coefficient: Service Category TCIs	Correlation Coefficient: Service Category RUIs
Composite	0.64	0.47

The indexed Total Resource Use measure has a high correlation to a risk adjusted composite utilization index, which was developed as a proxy to measure total resource consumption.

Prior submission: Please see Appendix A (page 15) for prior submission results.

In addition, the Total Resource Use measure was analyzed over time (2017 through 2019) to demonstrate stability and sensitivity to provider changes or improvement initiatives. Providers' performance across all three measures is relatively consistent across all three years and results are shown in the table below. The factors that drive variation between years within a provider are cost per unit and resource use management.

The results show that TCI has the most variation as it combines the changes for both price and resource use. The results also show that there is more variation in resource use over time than price. This indicates that providers are receiving similar price increases, but how providers are managing their patients' resource use is contributing more to the variation seen in costs.

Provider Group Size	TCI: 25th Percentile	TCI: Average	TCI: Median	TCI: 75th Percentile	Price: 25th Percentile	Price: Average	Price: Median	Price: 75th Percentile	RUI: 25th Percentile	RUI: Average	RUI: Median	RUI: 75th Percentile
Provider Group Size: <1,000	0.04	0.08	0.06	0.11	0.01	0.04	0.03	0.06	0.03	0.07	0.07	0.10
Provider Group Size: 1,000-2,000	0.03	0.07	0.05	0.09	0.02	0.05	0.04	0.06	0.04	0.07	0.07	0.08
Provider Group Size: 2,000+	0.01	0.04	0.03	0.05	0.01	0.02	0.02	0.03	0.01	0.03	0.03	0.05

Prior submission: Please see Appendix A (page 17) for prior submission results.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The Total Resource Use measure is valid as the critical data elements and the criteria applied produce a measure that accurately assesses various levels of performance. The norms in the measure are the network averages from the healthcare information derived from the MN market from included entities.

The Validity paper describes the results and conclusions from testing in detail: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-validity-analysis-2021-2021.pdf

In summary, the Total Resource Use measure accurately and consistently identified providers that are low or high performers with conclusions supported by known utilization measures.

There are high correlations between non-risk adjusted PMPM, ACG score and non-risk adjusted TCRRVs which indicate they are good measures of resources.

The ACGs, non-risk adjusted PMPMs, and non-risk adjusted TCRRVs have similar correlations to all utilization metrics which indicates the TCRRVs are performing as expected and are a solid measure of resources.

The indexed Total Resource Use measure scores have a moderately high correlation (0.47) to a risk adjusted composite utilization index score, which was developed as a proxy to measure total resource consumption.

The Total Resource Use measure differentiates between provider groups accurately as supported by the risk adjusted service utilization metrics.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

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

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

Prior submission: For this maintenance submission, there are no changes to HealthPartners Total Resource Use measure.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Details describing current exclusions can be found in the validity paper. There were no changes to measure criteria since re-endorsement in 2017.

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-validity-analysis-2021-2021.pdf

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* ***Note****:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The following exclusions and decision points remain unchanged from the original endorsed measures.

Nine month continuous enrollment – A nine month continuous enrollment was selected to balance business operations. Nine months allows for partial year enrollee. There was very little statistical difference in R-squared between six and twelve months.

Infants under age one are excluded due to slightly higher R-squared of the population without newborns, the required nine months enrollment criteria and variability in newborn costs, newborns under age one were excluded from the total cost of care measure.

Members over age 64 due are excluded due to potential incomplete claims data of Medicare eligible beneficiary.

The truncation level of 125,000 TCRRVs is used for the measure as the percent of members truncated has been stable over the last few years (0.3%-0.4%). Truncation levels are reviewed to stay aligned with healthcare costs and ensure stability of the measure.

Resource Use Measure Population Exclusion Funnel	Percent of Members	Percent of Total Resources
All Commercial Members	100%	100%
Members over 1	99%	97%
Members between 1-64	96%	89%
Members aged 1-64 and enrolled 9 months	77.6%	82.4%
Truncated at 125,000*	0.3%	77.1%
Member and Resources Included	77.6%	77.1%

Members are not removed from the measure

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☒ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

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. Information about the development of the grouper can be found here: http://acg.jhsph.org/; additionally, please refer to the ACG

Technical Reference Guide for supporting material: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/acg-technical-guide.pdf

ACG Grouper:

- Adjusted Clinical Groups (ACG System) were developed by Johns Hopkins University and allow comparisons between populations with varying illness burdens based on diagnoses, age and gender.
- Each unique member is assigned one of 93 ACG actuarial cells, which has a corresponding weight that reflects relative illness burden (e.g., relative expected resource consumption). Attributed members are assigned a risk score based on diagnoses on claims from the performance measurement period, as well as member age and gender

ACG-cell Risk Weights/Coefficients:

- The ACG risk weights measure relative resource variation between ACG actuarial cells/codes. Please see page 30-34 of the reference guide to view each ACG-cell risk weight. https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/acg-technical-guide.pdf
- Multiply each member's ACG weight by their eligible member months creating a total member ACG weight.

ACG Score:

- Each provider's attributed member ACG weights are summed to the provider level and divided by the sum of the attributed member months creating an ACG score for the provider.
- The provider's average ACG score is indexed to all attributed member's plan average ACG score.
- A member's total member ACG weight is updated to correspond with each year the Total Resource Use measure is measured.

Each of the 93 ACG actuarial cells can be considered a covariate of the multivariate risk model with the cell weights being the coefficients. The ACG cells are non-linear composites of the three risk factors: age, gender, diagnosis. Each member is assigned one of 93 covariates in the multivariate model and is based on the member's combination of age, gender and complete history of diagnosis codes.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable. All measures are clinically risk adjusted and limited to the commercial population.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

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. Information about the development of the grouper can be found here: http://acg.jhsph.org/; additionally, please refer to the ACG Technical Reference Guide for supporting material: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/acg-technical-guide.pdf

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. The 32 ADGs are listed on pages 4-6 in the reference guide, along with a table on pages 8-10 that provides guidance on how the five criteria are applied to each ADG.

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. The ACG assignment process can be found on page 12 of the reference guide.

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. Methodology and testing results can be found here: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-ses-social-risk-factors-technical-documentation-2020-2021.pdf

Also discuss any "ordering" of risk factor inclusion; for example, are social risk factors added after all clinical factors?

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.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

The risk factors included in ACG risk grouper were determined in the development of Johns Hopkins ACG risk grouper and are not available to the general public. The performance of the risk groupers are the basis for verifying the risk factors included in the model are sufficient to address clinic risk variation. The Society of Actuaries Accuracy of Claims-Based Risk Scoring Models (2016) findings also indicate the reliability and validity of the ACG risk grouper.

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

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g., prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

After risk adjusting for age, gender, and clinical risk, and limiting by insurance type, income does not significantly impact a patient's total resource use. As a potential practical use case example, the study also evaluated Resource Use provider group performance and found there was no discernible difference in performance when adjusting for income. The provider group analysis focused on the Resource Use measure to remove any bias based on price. The study considered two different data sources to study income variation, Census tract data and a commercially licensed data source available to HealthPartners with more specific income data.

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-ses-social-risk-factors-technical-documentation-2020-2021.pdf

The study utilized two independent data sources to evaluate income. The first was U.S. Census tracts. As defined by the U.S. Department of Commerce, "Census tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity that are updated by local participants prior to each decennial census as part of the Census Bureau's Participant Statistical Areas Program. The Census Bureau delineates census tracts in situations where no local participant existed or where state, local, or tribal governments declined to participate. The primary purpose of census tracts is to provide a stable set of geographic units for the presentation of statistical data.

Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people. A census tract usually covers a contiguous area; however, the spatial size of census tracts varies widely depending on the density of settlement. Census tract boundaries are delineated with the intention of being maintained over a long time so that statistical comparisons can be made from census to census. Census tracts occasionally are split due to population growth or merged as a result of substantial population decline."[5]

As noted, tracts estimate income by a general area and are not highly specific, introducing potential error and bias in the model.

HealthPartners utilized an additional data source to more accurately assess household income for purposes of this study. HealthPartners commercially licenses and has access to a large consumer database for other business purposes which gave us the ability to evaluate income with more specificity at the household level. Recognizing that it may not be feasible for all users to access a commercial database, HealthPartners pursued this deeper evaluation to more broadly understand the important question of whether or not to adjust cost performance measures by socioeconomic status independent of data availability. Household level income is derived using the midpoint of defined ranges of income by household (e.g., $20,000-$30,000) and capped at $250,000. Using the midpoint of a range introduces potential error in the evaluation whereas self-reported individual or household income would be most accurate.

Population-Based Evaluation

The evaluation tested the inclusion of income in addition to the factors already included in the measure specifications - age, gender, and clinical risk. Detailed measure criteria can be found in HealthPartners Technical Guidelines.

The study population included HealthPartners' full book of business of members, Commercial and Medicaid with TCOC criteria applied using services and claims generated throughout the 2019 time period. The study population included more than 555,000 members.

Three multiple linear regression models were created, each with one of the three metrics of interest as the dependent variable (total reimbursed amount per member per month, resource use per member per month, and price). Each model was identical in the use of income, ACG risk score, and insurance product (commercial vs Medicaid) as the independent variables. Resource use, reimbursed amount, price, and ACG scores were log transformed prior to developing the regression models to address the skewed nature of the data and adjust for heteroscedasticity. Insurance product was treated as a binary variable (commercial = 1, Medicaid = 0). The resulting coefficients were analyzed in terms of a 1% increase from average and their corresponding effect on the dependent variables.

Additionally, a model was created using only the endorsed measure criteria for the Resource Use measure (i.e., ACG and product only as the independent variables). The R^2 statistic from this model was compared against the R^2 statistic from the model that included income as an independent variable, allowing us to quantify the predictive value of income on resource use.

The same regression statistics and models were used with the second, more robust data source available to HealthPartners. This data contained more accurate income information, which was specific to household rather than tract, with household income defined using the midpoint or median of the income ranges. The more robust data source was available for 65% of HealthPartners' book of business members for 2019 and in the same proportions of commercial to Medicaid as in the previous evaluation.

Provider Group Performance Evaluation

A second evaluation was performed to provide a potential practical example of adjusting the TCOC and resource use measures by income using the Census and commercially licensed data sources. Resource Use Index was evaluated to remove known price variations between providers. HealthPartners' resource use results for its primary care network of commercial attributed members were used to evaluate provider group performance when adjusting for income. Medicaid was excluded from this evaluation as it has already been determined that provider performance results should be segmented by product.

There were 65 provider groups who met the measure criteria and were included in the evaluation using the Census tract data. The Total Resource Use measure is endorsed at a reliability level of 600 patients. Because the commercially licensed data source had available data for 65% of HealthPartners' book of business, there were 9 provider groups that failed to meet the 600 minimum and were excluded from the evaluation.

The variation between the average incomes using the Census tract data or the commercially licensed data source for each provider group was compared to the network average to adjust the provider's resource use index. It should be noted that while the adjustment can be made, the results should not be considered valid or reliable given the limitations inherent in each data source as described previously.

The regression analysis generated parameters that were translated into results based upon average cost, resource use, income, and ACG scores.

Table of Regression results using Census Tract Data

Model	1% Income Increase	1% ACG increase	Commercial vs. Medicaid Membership
Total Reimbursement	$ 0.09	$ 4.90	$ 207.41
Resource Use	$ 0.29	$ 5.00	$ (55.39)
Price	$ (0.18)	$ 0.05	$ 261.07

Model	R-Squared
Resource Use Endorsed Measure	0.5761
Resource Use Endorsed Measure + Income	0.5758

Using Census tract data, a 1% increase in income resulted in a $0.09 increase in total reimbursement, a $0.29 increase in resource use, and $0.18 decrease in price. The results highlight how significantly more the ACG score (clinical risk adjustment) and insurance product impact both the cost and resource use measures. For frame of reference, on average for the Midwest market, the total spend for a member per month (PMPM) is $500. The results of the evaluation show that a 1% increase in risk score accounts for a $4.90 or roughly 1% increase in PMPM.

Product also contributed significantly with there being a $207 dollar difference in cost between commercial and Medicaid. The variation in resource use was much less, however, still significant with Medicaid covered members utilizing $55 more dollars of resources. The fact that Medicaid's cost per service is approximately half that of commercial rates drives the differences between the TCOC and Resource Use results. The R^2 results further emphasize that ACG score and insurance type are the main drivers of cost and resource use variation and income does not provide any additional predictive power.

Table of Regression results using Commercially Licensed Data Source

Model	1% Income Increase	1% ACG increase	Commercial vs. Medicaid Membership
Total Reimbursement	$ 0.08	$ 5.32	$ 213.93
Resource Use	$ 0.13	$ 5.40	$ (69.04)
Price	$ (0.05)	$ 0.05	$ 281.22

Model	R-Squared
Resource Use Endorsed Measure	0.57114
Resource Use Endorsed Measure + Income	0.57132

Using the commercially purchased data source, with income by household, a 1% increase in income resulted in an $0.08 increase for total reimbursement, $0.13 increase in resource use, and $0.05 decrease in price. This is telling, as when using a data source that is more specific, income is even less impactful on TCOC and resource use while ACG and product type show similar results.

Results– Provider Group Performance Evaluation

Provider group performance of the Resource Use measure was evaluated to test the impact of income adjustment on the Resource Use measure. Provider group results for both data sources, Census tract and commercially licensed, are shown below using HealthPartners' commercial provider network. The Resource Use Index (RUI) is calculated using the endorsed measure criteria. The second RUI is calculated using the endorsed measure criteria ***with income adjustment.***

The Census tract data evaluated 65 provider groups and the commercially licensed data source evaluated 56 provider groups. Because the population of patients used between the two data sources is different, Provider Group 01 in the Census tract chart is not the same as Provider Group 01 in the commercially licensed chart. Provider group numbers in the Census tract chart are numbered based on ascending Total Cost Index found in the appendix of the study paper. Provider groups for both charts are sorted in ascending order using the RUI.

On average there was less than a 1% change in performance for provider groups when income was introduced into the model for the Resource Use measure when using Census tract data. This impact was reduced on average to less than a 0.50% when using the commercially licensed data source with more specific income data. Considering the Resource Use measure identifies provider performance levels (indices) that span greater than 137% as identified below, the less than 1% adjustment was considered insignificant when comparing provider performance. Provider Group charts begin on the following page.

Census Tract Data Source

RUI	Data
RUI Min	0.86
RUI Max	1.18
RUI Max/Min % Difference	137%
Average % change with income adjustment	0.91%

Commercially Licensed Data Source

RUI	Data
RUI Min	0.86
RUI Max	1.20
RUI Max/Min % Difference	140%
Average % change with income adjustment	0.47%

Provider Group Detailed Results – Census Data

Provider Group	TCI	RUI (endorsed measure)	RUI (endorsed measure + income)	Average Income	RUI Change	Pct RUI Change
Provider 58	1.30	0.86	0.87	$ 51,182.66	0.0125	1.45%
Provider 01	0.80	0.87	0.88	$ 60,871.13	0.0096	1.09%
Provider 14	0.95	0.89	0.90	$ 51,196.01	0.0058	0.65%
Provider 42	1.07	0.91	0.92	$ 57,184.20	0.0115	1.26%
Provider 07	0.91	0.92	0.93	$ 53,262.19	0.0155	1.69%
Provider 38	1.05	0.94	0.95	$ 52,994.97	0.0118	1.26%
Provider 51	1.17	0.94	0.95	$ 48,573.69	0.0081	0.87%
Provider 49	1.13	0.94	0.95	$ 54,522.47	0.0134	1.43%
Provider 40	1.05	0.94	0.95	$ 55,143.95	0.0133	1.41%
Provider 09	0.92	0.94	0.94	$ 49,821.34	-0.0075	-0.79%
Provider 50	1.17	0.95	0.96	$ 74,230.68	0.0196	2.07%
Provider 10	0.92	0.96	0.96	$ 54,236.28	0.0016	0.16%
Provider 52	1.19	0.96	0.97	$ 59,432.45	0.0121	1.26%
Provider 12	0.92	0.96	0.97	$ 57,038.75	0.0076	0.79%
Provider 63	1.60	0.96	0.97	$ 59,923.30	0.0112	1.17%
Provider 46	1.13	0.97	0.98	$ 56,657.27	0.0127	1.31%
Provider 60	1.50	0.97	0.98	$ 46,884.50	0.0109	1.12%
Provider 59	1.45	0.98	0.99	$ 74,671.53	0.0091	0.93%
Provider 06	0.91	0.98	0.98	$ 50,511.60	-0.0015	-0.16%

Provider Group	TCI	RUI (endorsed measure)	RUI (endorsed measure + income)	Average Income	RUI Change	Pct RUI Change
Provider 08	0.92	0.98	0.98	$ 51,481.13	-0.0078	-0.80%
Provider 37	1.04	0.98	1.00	$ 63,017.62	0.0144	1.46%
Provider 21	0.98	0.99	1.00	$ 85,046.08	0.0112	1.14%
Provider 03	0.84	0.99	1.00	$ 75,988.94	0.0065	0.66%
Provider 64	1.61	0.99	1.01	$ 53,580.68	0.0178	1.80%
Provider 16	0.96	0.99	1.00	$ 60,513.25	0.0128	1.29%
Provider 35	1.04	0.99	1.00	$ 56,581.35	0.0122	1.23%
Provider 02	0.81	1.00	1.00	$ 53,033.63	-0.0015	-0.15%
Provider 17	0.96	1.00	0.98	$ 78,737.12	-0.0150	-1.50%
Provider 62	1.56	1.00	1.00	$ 68,995.93	0.0034	0.34%
Provider 20	0.98	1.01	1.02	$ 63,162.88	0.0142	1.41%
Provider 18	0.97	1.01	1.01	$ 87,449.16	-0.0049	-0.48%
Provider 05	0.90	1.01	1.02	$ 75,724.77	0.0101	1.00%
Provider 47	1.13	1.02	1.02	$ 81,800.09	0.0081	0.80%
Provider 19	0.97	1.02	1.03	$ 80,344.27	0.0093	0.92%
Provider 15	0.96	1.02	1.01	$ 68,200.83	-0.0083	-0.81%
Provider 39	1.05	1.02	1.02	$ 45,478.38	-0.0031	-0.30%
Provider 04	0.90	1.02	1.03	$ 55,181.82	0.0096	0.94%
Provider 56	1.28	1.02	1.04	$ 85,397.81	0.0128	1.25%
Provider 11	0.92	1.03	1.02	$ 79,078.76	-0.0041	-0.40%
Provider 26	1.01	1.03	1.04	$ 76,886.96	0.0078	0.76%
Provider 43	1.07	1.03	1.04	$ 58,557.65	0.0136	1.32%
Provider 23	0.99	1.03	1.04	$ 76,364.65	0.0073	0.71%
Provider 45	1.11	1.03	1.04	$ 51,695.82	0.0084	0.81%
Provider 48	1.13	1.03	1.04	$ 80,133.18	0.0079	0.77%
Provider 29	1.02	1.04	1.02	$ 62,718.98	-0.0161	-1.56%
Provider 28	1.01	1.04	1.04	$ 76,650.16	-0.0027	-0.26%
Provider 32	1.03	1.05	1.05	$ 57,718.34	-0.0021	-0.20%
Provider 33	1.03	1.05	1.05	$ 60,952.95	-0.0019	-0.18%

Provider Group	TCI	RUI (endorsed measure)	RUI (endorsed measure + income)	Average Income	RUI Change	Pct RUI Change
Provider 22	0.98	1.06	1.05	$ 56,343.84	-0.0100	-0.95%
Provider 57	1.30	1.06	1.08	$ 73,585.43	0.0148	1.39%
Provider 41	1.07	1.06	1.06	$ 61,287.61	-0.0058	-0.54%
Provider 13	0.93	1.07	1.07	$ 88,286.87	0.0033	0.31%
Provider 36	1.04	1.07	1.06	$ 76,082.19	-0.0087	-0.82%
Provider 54	1.20	1.08	1.08	$ 80,419.35	-0.0051	-0.47%
Provider 31	1.03	1.09	1.08	$ 55,164.25	-0.0071	-0.65%
Provider 25	1.00	1.10	1.08	$ 55,820.84	-0.0123	-1.12%
Provider 53	1.19	1.10	1.11	$ 60,338.84	0.0112	1.02%
Provider 61	1.54	1.10	1.12	$ 42,557.01	0.0168	1.52%
Provider 27	1.01	1.12	1.11	$ 94,343.76	-0.0014	-0.13%
Provider 34	1.04	1.12	1.11	$ 52,722.55	-0.0089	-0.79%
Provider 44	1.10	1.13	1.14	$ 85,490.55	0.0112	0.99%
Provider 55	1.21	1.13	1.14	$ 86,685.54	0.0089	0.79%
Provider 24	0.99	1.13	1.12	$ 82,723.92	-0.0081	-0.71%
Provider 30	1.03	1.13	1.12	$ 89,318.28	-0.0082	-0.72%
Provider 65	1.94	1.18	1.18	$ 85,086.95	-0.0049	-0.41%
Provider 66	2.03	1.39	1.38	$ 75,167.49	-0.0016	-0.12%

Provider Group Detailed Results – Commercially Licensed Data

Provider Group	TCI	RUI (endorsed measure)	RUI (endorsed measure + income)	Average Income	RUI Change	Pct RUI Change
Provider 01	0.78	0.86	0.87	$ 86,036.29	0.0067	0.78%
Provider 02	1.06	0.91	0.92	$ 87,569.64	0.0062	0.68%
Provider 03	1.18	0.93	0.93	$ 88,472.12	0.0060	0.64%
Provider 04	1.21	0.93	0.93	$ 98,563.70	0.0030	0.32%

Provider Group	TCI	RUI (endorsed measure)	RUI (endorsed measure + income)	Average Income	RUI Change	Pct RUI Change
Provider 05	1.13	0.93	0.93	$ 94,120.09	0.0043	0.46%
Provider 06	1.10	0.93	0.93	$ 91,380.65	0.0051	0.55%
Provider 07	0.99	0.94	0.94	$ 85,467.71	0.0069	0.73%
Provider 08	1.02	0.94	0.94	$ 95,145.14	0.0040	0.42%
Provider 09	0.89	0.94	0.94	$ 120,878.09	-0.0037	-0.39%
Provider 10	1.22	0.94	0.95	$ 88,289.82	0.0060	0.64%
Provider 11	0.93	0.96	0.96	$ 96,516.24	0.0036	0.37%
Provider 12	1.26	0.96	0.97	$ 83,907.02	0.0073	0.76%
Provider 13	0.89	0.96	0.97	$ 101,396.05	0.0021	0.22%
Provider 14	1.01	0.97	0.98	$ 89,304.06	0.0057	0.59%
Provider 15	1.19	0.98	0.98	$ 97,706.43	0.0032	0.33%
Provider 16	0.87	0.98	0.99	$ 98,260.28	0.0030	0.31%
Provider 17	1.03	0.99	0.99	$ 97,811.08	0.0032	0.32%
Provider 18	1.03	0.99	1.00	$ 93,852.88	0.0044	0.44%
Provider 19	0.99	0.99	1.00	$ 85,794.75	0.0068	0.68%
Provider 20	1.59	0.99	1.00	$ 89,623.34	0.0056	0.57%
Provider 21	0.85	1.00	1.00	$ 110,798.85	-0.0007	-0.07%
Provider 22	1.17	1.00	1.01	$ 95,487.43	0.0039	0.39%
Provider 23	1.02	1.00	1.01	$ 85,850.93	0.0067	0.67%
Provider 24	0.99	1.01	1.00	$ 122,459.73	-0.0042	-0.41%
Provider 25	1.03	1.01	1.01	$ 101,049.14	0.0022	0.22%
Provider 26	0.93	1.01	1.01	$ 106,245.91	0.0007	0.07%
Provider 27	1.05	1.02	1.02	$ 113,982.13	-0.0016	-0.16%
Provider 28	0.99	1.02	1.02	$ 122,997.47	-0.0043	-0.42%
Provider 29	1.18	1.02	1.03	$ 98,794.22	0.0029	0.28%
Provider 30	1.70	1.02	1.03	$ 79,617.81	0.0086	0.84%
Provider 31	0.97	1.03	1.02	$ 135,777.93	-0.0081	-0.79%
Provider 32	1.11	1.03	1.04	$ 87,234.29	0.0063	0.61%
Provider 33	1.05	1.04	1.02	$ 155,630.21	-0.0141	-1.35%

Provider Group	TCI	RUI (endorsed measure)	RUI (endorsed measure + income)	Average Income	RUI Change	Pct RUI Change
Provider 34	1.67	1.04	1.04	$ 100,797.03	0.0023	0.22%
Provider 35	1.06	1.04	1.04	$ 119,035.76	-0.0031	-0.30%
Provider 36	0.99	1.04	1.03	$ 143,564.91	-0.0105	-1.00%
Provider 37	1.03	1.05	1.04	$ 137,618.95	-0.0087	-0.83%
Provider 38	1.07	1.05	1.05	$ 92,556.75	0.0047	0.45%
Provider 39	1.05	1.05	1.05	$ 111,157.12	-0.0008	-0.08%
Provider 40	1.63	1.05	1.05	$ 104,873.45	0.0011	0.10%
Provider 41	1.05	1.06	1.06	$ 111,366.04	-0.0009	-0.08%
Provider 42	1.29	1.06	1.07	$ 86,786.91	0.0065	0.61%
Provider 43	1.23	1.07	1.07	$ 121,562.20	-0.0039	-0.36%
Provider 44	0.93	1.07	1.07	$ 110,460.23	-0.0006	-0.05%
Provider 45	1.00	1.08	1.07	$ 118,541.16	-0.0030	-0.28%
Provider 46	1.00	1.08	1.07	$ 128,348.85	-0.0059	-0.55%
Provider 47	1.13	1.08	1.08	$ 116,821.29	-0.0025	-0.23%
Provider 48	1.02	1.09	1.09	$ 113,478.74	-0.0015	-0.14%
Provider 49	1.23	1.10	1.11	$ 87,181.46	0.0063	0.58%
Provider 50	1.04	1.11	1.10	$ 134,920.71	-0.0079	-0.71%
Provider 51	1.03	1.13	1.12	$ 129,914.90	-0.0064	-0.57%
Provider 52	1.25	1.13	1.14	$ 88,340.73	0.0060	0.53%
Provider 53	1.08	1.14	1.13	$ 123,463.42	-0.0045	-0.39%
Provider 54	1.02	1.20	1.21	$ 84,086.96	0.0073	0.60%
Provider 55	1.36	1.20	1.21	$ 92,921.72	0.0046	0.39%
Provider 56	2.06	1.22	1.21	$ 135,014.93	-0.0079	-0.65%

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

Correlations and regression analysis utilized in both validity and the socioeconomic testing papers as well as the results in the Society of Actuaries study indicate that the statistical model used to adjust cost variation is effective. Additionally, because the commercial population's use of the healthcare system is so significantly different from the Medicaid and Medicare populations, through the benefits covered, the predominant conditions treated, and the prices of the services rendered, segmentation is required.

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

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. An evaluation between commercial and Medicaid covered members was also conducted in the socioeconomic testing, highlighting the variation in resource use (results included in 2b3.9.).

Metric	R-Squared: ACG	R-Squared: Non-Risk Adj PMPMs
Non-Risk Adj PMPM	0.45	1.00
Non-Risk Adj TCRRVs	0.90	0.55
ACG Risk Adj TCI	0.00	0.52
ACG Risk Adj RUI	0.01	0.07
Price	0.00	0.36

The non-risk adjusted Total Cost Relative Resource Values coefficient of 0.95 indicates a high correlation between total resource use and risk score.

Metric	Correlation Coefficient: ACG	Correlation Coefficient: Non-Risk Adj PMPMs
Non-Risk Adj PMPM	0.67	1.00
Non-Risk Adj TCRRVs	0.95	0.74
ACG Risk Adj TCI	0.00	0.72
ACG Risk Adj RUI	0.11	0.26
Price	-0.06	0.60

Validity Paper (see page 5):

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-validity-analysis-2021-2021.pdf

Socioeconomic Testing Paper (see page 4): https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-ses-social-risk-factors-technical-documentation-2020-2021.pdf

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

Detailed results can be found on page 4 and 5 of the socioeconomic testing paper:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-ses-social-risk-factors-technical-documentation-2020-2021.pdf

Using Census Tract Data, the stratification results are shown in the far right column.

Model	1% Income Increase	1% ACG increase	Commercial vs. Medicaid Membership
Total Reimbursement	$ 0.09	$ 4.90	$ 207.41
Resource Use	$ 0.29	$ 5.00	$ (55.39)
Price	$ (0.18)	$ 0.05	$ 261.07

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Detailed results can be found on page 4 and 5 of the socioeconomic testing paper:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/tcoc-resource-use-ses-social-risk-factors-technical-documentation-2020-2021.pdf

Product contributed significantly with there being a $207 dollar difference in cost between commercial and Medicaid. The variation in resource use was much less, however, still significant with Medicaid covered members utilizing $55 more dollars of resources. The fact that Medicaid's cost per service is approximately half that of commercial rates drives the differences between the TCOC and Resource Use results.

2b3.11. Optional Additional Testing for Risk Adjustment (***not required****, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Performance is measured on an Index basis relative to 1.00 where each one point (0.01) variation from 1.00 (average) represents a 1% deviation from average. Statistical significance ranges of performance are not necessary as the measure is based on the full population. The results can be analyzed by percentile, percent from mean, standard deviation and clustering methods, this is dependent upon the business application of the measure.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Provider Group	Average ACG Score: 2017	Average ACG Score: 2018	Average ACG Score: 2019	TCI: 2017	TCI: 2018	TCI: 2019	Price Index: 2017	Price Index: 2018	Price Index: 2019	Resource Use Index: 2017	Resource Use Index: 2018	Resource Use Index: 2019
Provider 01	1.12	1.16	1.17	0.86	0.80	0.80	0.89	0.89	0.91	0.94	0.88	0.87
Provider 02	1.23	1.27	1.18	0.83	0.78	0.81	0.79	0.79	0.81	1.03	0.97	1.00
Provider 03	1.07	1.07	1.02	0.81	0.81	0.84	0.85	0.86	0.85	0.93	0.93	0.99
Provider 04	1.26	1.21	1.14	0.92	0.79	0.90	0.90	0.89	0.88	1.00	0.88	1.02
Provider 05	1.33	1.38	1.27	0.79	0.78	0.90	0.83	0.89	0.89	0.93	0.86	1.01
Provider 06	0.86	0.87	0.84	0.91	0.90	0.91	0.92	0.94	0.92	0.97	0.94	0.98
Provider 07	1.04	1.00	0.98	1.02	1.04	0.91	1.00	1.02	0.99	1.00	1.01	0.92
Provider 08	0.53	0.53	0.50	0.97	0.99	0.92	0.89	0.90	0.93	1.06	1.08	0.98
Provider 09	1.13	1.12	1.11	0.92	0.92	0.92	0.95	0.96	0.97	0.94	0.94	0.94

Provider Group	Average ACG Score: 2017	Average ACG Score: 2018	Average ACG Score: 2019	TCI: 2017	TCI: 2018	TCI: 2019	Price Index: 2017	Price Index: 2018	Price Index: 2019	Resource Use Index: 2017	Resource Use Index: 2018	Resource Use Index: 2019
Provider 10	1.02	1.03	1.05	0.95	0.95	0.92	0.92	0.94	0.96	1.00	0.99	0.96
Provider 11	1.11	1.10	1.00	0.79	0.90	0.92	0.89	0.89	0.89	0.87	1.00	1.03
Provider 12	0.94	0.99	1.03	0.98	0.94	0.92	0.93	0.94	0.96	1.03	0.98	0.96
Provider 13	0.60	0.54	0.61	0.97	0.97	0.93	0.87	0.84	0.87	1.08	1.13	1.07
Provider 14	0.96	1.00	1.03	1.10	1.11	0.95	1.12	1.11	1.06	0.96	0.98	0.89
Provider 15	1.07	1.08	1.13	0.98	0.94	0.96	0.92	0.94	0.94	1.04	0.99	1.02
Provider 16	0.86	0.87	0.90	1.01	0.94	0.96	0.92	0.93	0.97	1.06	0.99	0.99
Provider 17	0.61	0.59	0.58	0.94	0.91	0.96	0.95	0.93	0.96	0.96	0.95	1.00
Provider 18	1.02	1.02	1.03	0.95	0.99	0.97	0.92	0.94	0.96	1.01	1.03	1.01
Provider 19	NA	NA	0.85	NA	NA	0.97	NA	NA	0.96	NA	NA	1.02
Provider 20	0.90	0.99	1.01	0.95	0.92	0.98	0.92	0.93	0.97	1.01	0.97	1.01
Provider 21	0.95	0.92	0.89	1.08	0.98	0.98	1.01	0.97	0.99	1.04	1.00	0.99
Provider 22	1.12	1.09	1.07	1.00	1.02	0.98	0.90	0.90	0.93	1.09	1.11	1.06
Provider 23	1.34	1.27	1.28	0.98	0.86	0.99	0.96	0.94	0.96	0.99	0.90	1.03
Provider 24	NA	1.29	1.31	NA	1.05	0.99	NA	0.87	0.88	NA	1.19	1.13
Provider 25	1.01	1.01	1.04	0.90	0.95	1.00	0.88	0.91	0.91	0.99	1.02	1.10
Provider 26	0.92	0.96	0.97	1.01	1.00	1.01	0.99	0.95	0.98	1.00	1.03	1.03
Provider 27	1.07	1.06	1.02	0.97	0.95	1.01	0.89	0.90	0.90	1.05	1.04	1.12
Provider 28	0.82	0.83	0.84	1.00	0.95	1.01	0.95	0.95	0.97	1.02	0.99	1.04
Provider 29	0.54	0.57	0.56	1.00	1.00	1.02	0.93	0.97	0.98	1.04	1.02	1.04
Provider 30	1.11	1.10	1.10	1.00	1.04	1.03	0.90	0.90	0.91	1.09	1.14	1.13
Provider 31	0.88	0.83	0.83	0.95	1.00	1.03	0.90	0.95	0.95	1.02	1.04	1.09
Provider 32	1.09	1.07	1.08	1.01	1.03	1.03	0.95	0.96	0.99	1.04	1.05	1.05
Provider 33	1.03	1.03	1.05	1.00	1.00	1.03	0.96	0.96	0.98	1.02	1.02	1.05

The red line divides providers between above and below the average total cost index (1.00).

Provider Group	Average ACG Score: 2017	Average ACG Score	Average ACG Score	TCI: 2017	TCI: 2018	TCI: 2019	Price Index: 2017	Price Index: 2018	Price Index: 2019	Resource Use Index: 2017	Resource Use Index: 2018	Resource Use Index: 2019
Provider 34	0.98	0.99	0.98	1.01	1.05	1.04	0.90	0.91	0.92	1.09	1.12	1.12
Provider 35	0.90	0.94	0.92	0.98	0.99	1.04	1.01	1.08	1.05	0.95	0.90	0.99
Provider 36	0.53	0.53	0.52	1.03	1.02	1.04	0.94	0.96	0.97	1.07	1.05	1.07
Provider 37	1.06	1.01	1.09	0.94	1.04	1.04	1.00	1.04	1.06	0.92	0.98	0.98
Provider 38	0.94	0.95	0.92	1.03	1.06	1.05	1.06	1.14	1.12	0.95	0.91	0.94

Provider Group	Average ACG Score: 2017	Average ACG Score	Average ACG Score	TCI: 2017	TCI: 2018	TCI: 2019	Price Index: 2017	Price Index: 2018	Price Index: 2019	Resource Use Index: 2017	Resource Use Index: 2018	Resource Use Index: 2019
Provider 39	1.11	1.11	1.10	1.04	1.04	1.05	1.00	1.02	1.03	1.01	1.00	1.02
Provider 40	NA	NA	1.39	NA	NA	1.05	NA	NA	1.12	NA	NA	0.94
Provider 41	0.95	0.97	1.02	0.99	0.94	1.07	0.93	0.94	1.00	1.04	0.99	1.06
Provider 42	0.90	0.86	0.89	1.22	1.18	1.07	1.14	1.15	1.17	1.05	1.01	0.91
Provider 43	0.92	0.91	0.95	1.19	1.10	1.07	1.09	1.04	1.04	1.06	1.04	1.03
Provider 44	1.06	1.05	1.09	0.93	1.19	1.10	0.92	1.01	0.97	0.98	1.16	1.13
Provider 45	0.86	0.93	0.96	1.23	1.21	1.11	1.10	1.15	1.08	1.09	1.04	1.03
Provider 46	0.91	0.90	0.89	1.11	1.18	1.13	1.18	1.18	1.16	0.92	0.98	0.97
Provider 47	NA	0.98	0.95	NA	1.02	1.13	NA	1.11	1.11	NA	0.90	1.02
Provider 48	NA	0.82	0.83	NA	1.17	1.13	NA	1.06	1.09	NA	1.08	1.03
Provider 49	0.84	0.85	0.90	1.21	1.20	1.13	1.20	1.19	1.21	0.98	0.99	0.94
Provider 50	0.93	0.84	0.89	1.37	1.26	1.17	1.24	1.22	1.24	1.08	1.02	0.95
Provider 51	0.93	0.85	0.89	1.20	1.21	1.17	1.19	1.24	1.25	0.98	0.96	0.94
Provider 52	1.09	1.00	0.93	1.46	1.26	1.19	1.44	1.39	1.23	0.99	0.89	0.96
Provider 53	1.06	1.10	1.08	1.17	1.14	1.19	1.07	1.09	1.08	1.06	1.03	1.10
Provider 54	0.82	0.78	0.78	1.22	1.21	1.20	1.09	1.07	1.11	1.09	1.11	1.08
Provider 55	0.88	0.94	0.84	1.12	1.15	1.21	1.19	1.06	1.07	0.91	1.07	1.13
Provider 56	1.02	1.06	1.01	1.25	1.22	1.28	1.19	1.24	1.25	1.02	0.97	1.02
Provider 57	NA	0.89	0.86	NA	1.29	1.30	NA	1.28	1.22	NA	0.99	1.06
Provider 58	0.76	0.79	0.84	1.40	1.45	1.30	1.47	1.47	1.50	0.93	0.96	0.86
Provider 59	0.83	0.80	0.86	1.34	1.48	1.45	1.26	1.33	1.48	1.04	1.10	0.98
Provider 60	1.03	0.94	0.96	1.42	1.52	1.50	1.44	1.51	1.55	0.96	0.98	0.97
Provider 61	0.83	0.78	0.86	1.27	1.49	1.54	1.21	1.30	1.40	1.02	1.13	1.10
Provider 62	0.81	0.79	0.76	1.65	1.41	1.56	1.46	1.45	1.56	1.10	0.95	1.00
Provider 63	0.97	1.01	1.02	1.66	1.68	1.60	1.59	1.75	1.67	1.02	0.94	0.96
Provider 64	0.90	0.92	0.83	1.60	1.63	1.61	1.62	1.68	1.63	0.96	0.96	0.99
Provider 65	1.41	1.48	1.52	2.09	1.99	1.94	1.51	1.58	1.64	1.35	1.24	1.18
Network Total	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities?
(i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The Total Resource Use measure can effectively identify variation in performance levels.

Practically meaningful difference in performance will vary by use of the measures. This is because some uses may have a higher threshold for differences. For example, a 10% difference in performance when the result is used for public reporting could be very meaningful in terms of provider patient growth and retention strategies. The same 10% difference may not be as meaningful when using the measures internally for improvement work and identification of a work plan.

The following will give a general sense of the dispersion of the scoring:

Out of the 65 provider groups measured in Total Resource Use measure:

- 26 were better than average
- 3 were 10% better than average
- 12 were 10% higher than average
- 50 were within 10% of the average

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

This is a full population-based measure, all data is included in the measure.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; **if no empirical sensitivity analysis**, identify the approaches for handling missing data that were considered and pros and cons of each*)

This is a full population-based measure, all data is included in the measure.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; **if no empirical analysis**, provide rationale for the selected approach for missing data*)

This is a full population-based measure, all data is included in the measure.

Feasibility

F.1. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

F.1.1. Data Elements Generated as Byproduct of Care Processes.

Other

If other: Health Plan Claims data system

F.2. Electronic Sources

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

F.2.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*)

ALL data elements are in defined fields in a combination of electronic sources

F.2.1a. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.

F.2.2. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.

Attachment:

F.3. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

F.3.1. Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

Since endorsement we have received some general feedback regarding implementation of the measure. This has helped shape some of the materials and additional testing we've conducted since the measures were first released. HealthPartners has organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download necessary tools to run the measure, free of charge. In addition, HealthPartners uses SAS to run the measure and not every organization has or uses this software. To address this, HealthPartners organized non-SAS user instructions. By creating these resources and software and putting them in the public domain it has resulted in expanded use. A few users have successfully implemented the NQF-endorsed Total Resource Use measure according to the specifications, however they are using their previously purchased risk grouper (not ACG).

F.3.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, and algorithm)?

The measure and software are available free of charge at www.healthpartners.com/tcoc;

The TCOC measure download options are available at:
https://www.healthpartners.com/hp/about/tcoc/toolkit/index.html

The ACG System is widely available within the public and private sectors in the US and abroad.(Bibliography: http://acg.jhsph.org/index.php/resource-center-83/acg-bibliography) The pricing for the ACG System varies for commercial and government entities but is generally based on a per member per year license that is tiered to provide lower per member costs for larger entities. For a commercial plan there is a base fee of $33,000 annually with incremental costs between $0.05 and $0.41 per member per year based on volume, which is inclusive of both license fees and support. Discounted fees are available for government entities and other grant funded not-for-profit entities. Additionally, Johns Hopkins offers research licenses for a very modest cost for academic users incorporating ACGs into published research:
http://www.acg.jhsph.org/index.php?option=com_content&view=article&id=137&Itemid=94

The ACG System technical guide is available here: https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/acg-technical-guide.pdf

F.3.3. If there are any fees associated with the use of this measure as specified, attach the fee schedule here. (Save file as: F3_3_FeeSchedule)

FeeScheduleTemplate_Proprietary_Fees_V2.0SubmissionForm-Johns_Hopkins_ACG_System_2021-04-637541805703764558.xlsx

Usability and Use

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

U.1.1. Current and Planned Use

Specific Plan for Use	Current Use (for current use provide URL)
Specific Plan	Public Reporting See URL https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf Public Health/Disease Surveillance See URL https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf Payment Program See URL https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf Quality Improvement (external benchmarking to organizations) See URL https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf Quality Improvement (Internal to the specific organization) See URL https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf

U.1.2. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Since endorsement in 2012, uptake of the Total Resource Use measure has expanded across 42 states in the country and used by both national and regional organizations (Coverage). The measure has been used in accountability applications and publicly reported in multiple states for driving improvement.
The following link highlights organizations across the country that have adopted the measure and are currently using it for at least one of the uses noted above, including some crossover of multiple uses for some organizations. The URLs of the specific programs are included within the link below to appropriately capture each organization's purpose described in their own words.
Because some of the organizations are using the measure for multiple uses, we have included them based on their predominant category.
https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf
Public Reporting
HealthPartners – Public Reporting, Payment Program, Quality Improvement with Benchmarking

- As a health plan, HealthPartners uses the Total Resource Use measure to incentivize providers to meet Triple Aim goals, optimizing health and patient experience while improving affordability. HealthPartners publicly reports provider group cost results for purposes of transparency for employers, providers, and consumers. The resource use results are paired with Total Cost of Care, quality and experience metrics to promote quality improvement with benchmarking across providers.
- HealthPartners has shared savings payment agreements with over 85% of its primary care providers which has increased provider engagement and sharing of appropriate risk as a partnership to lower cost for providers and patients while maintaining quality and experience. Additionally, in conjunction with the Total Cost of Care measure, HealthPartners has begun building upon it by implementing new payment reform models that align incentives among specialists and hospitals to support shared savings with primary care. The new methods include bundled payments for episodes of care as well as models that move away from fee for service and promote coordination of care.

MN Community Measurement – Public Reporting, Quality Improvement with Benchmarking

- Beginning in 2016, MNCM publicly reported Total Resource Use data by provider group in Minnesota using HealthPartners endorsed Total Resource Use measure. Through their multi-stakeholder collaborative process, they were able to collect cost data from four health plans and publicly spread the use of the measure to all provider groups in Minnesota, promoting transparency.

Network for Regional Healthcare Improvement – Public Reporting, Quality Improvement with Benchmarking, Quality Improvement

- Eleven regions were part of a project to develop a standardized method of reporting total cost and resource use by using the HealthPartners endorsed measures. During 2015, seven regions were able to share healthcare cost information on over 5 million patients attributed to 20,000 individual physicians through practice level reports. Their work is described in detail in the provided link.

Payment Program

The Alliance

- Utilizes the measure for provider contracting and incentives. To understand high value primary care providers and employer costs, to evaluate price efficiencies. The measures inform decisions and influence contract negotiations with providers. The measure provides invaluable standardization to measure overall value.

Public Health/Disease Surveillance

The University of Iowa Public Policy Center

- Research evaluation for assessing state health system transformation under the State Innovation Model initiative. Utilize the measure to establish overall cost baseline numbers and comparison over time. Currently data is shared with the state to inform possible future initiatives with the community.

Delaware Health Information Network (third party of Delaware state agency)

- Total Cost of Care and Total Resource Use measures are currently being reported for the state of Delaware using Delaware´s Health Care Claims Database (APCD) administrated by DHIN.

Quality Improvement with Benchmarking

Maine Health Management Coalition – regional collaborative

- Commercial premium costs will be measured against benchmarks using the TCOC and Resource Use measures with plans for future public reporting.

Oregon Health Care Quality Corporation – regional collaborative

- In 2015, Q Corp released Clinic Comparison Reports featuring cost, utilization and quality measures to over 150 primary care clinics in Oregon.

HealthInsight and Utah Department of Health, Washington Health Alliance – regional collaboratives

- Regional collaboratives participating in the Network for Regional Healthcare Improvement's project to develop a standardized method of reporting total cost and resource use.

Center for Improving Value in Health Care (CIVHC) – regional collaborative

- Recently began providing results to Colorado primary care groups to help them see how their practice patterns compare.

Midwest Health Initiative – regional collaborative

- Shares data with physician groups and practice sites through community reports with future plans for public reporting.

Quality Improvement

Provider Groups in Minnesota, Wisconsin, and Iowa

- Having payment agreements with HealthPartners, several provider groups see the value and are actively engaged in utilizing the Total Resource Use measure. They shared with us how they are using the measure within their own practice and their letters of support are included in the link.

Priority Health

- Evaluating practice efficiencies and pricing fluctuations across Accountable Care Networks.

Providence Health Plans

- Provide efficiency profiling and increasing engagement for improvement and better referral decision making.

Onpoint Health Data

- State organization are utilizing the data for program evaluation.

U.1.3. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not applicable

U.1.4. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not applicable

U.2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation. How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

HealthPartners has been measuring total resource use with its provider network since 1995. Since endorsement in 2012, uptake of the Total Resource Use measure has expanded across 42 states in the country and used by both national and regional organizations. The utility continues to expand across measured entities. As described under our attribution logic, the measure can be applied by geographic region or defined population. The measure has been used in accountability applications and publicly reported in multiple states for driving improvement.

HealthPartners produces a suite of comprehensive assessment reports, providing depth under the total cost of care and resource use measures, on a quarterly basis to help providers pinpoint opportunities for improvement. The suite of reports provide comparative benchmarking information for providers to assess the effectiveness of their care model. This information identifies where the provider deviates from benchmarks in cost, resource use, admissions, surgeries, ED visits, high tech radiology, primary care and specialty visits and pharmacy usage. This information is drillable to the condition (e.g., diabetes, back pain, mental health) and procedure level.

Through ongoing in-person or virtual meetings we communicate and consult with providers on identifying opportunities. The opportunities identified can be turned into action through patient registries or applications

that we also share monthly with providers which include patient information that can be incorporated into a provider's workflow to support improved point of care service. The patient information includes quality and conditions that need to be addressed, predictive analytics showcasing risk of inpatient admission and high future costs, quality and conditions that need to be addressed as well as utilization markers (e.g., admissions, ED visits, surgeries, conditions, gaps in addressing quality and chronic conditions, specialty visits specialists they've seen). Providers can then use this information to help guide outreach opportunities and pre-visit planning with patients. We encourage and maintain open communication with providers on maximizing the use of these reports for patient management throughout the year.

U.2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

HealthPartners provides quarterly comprehensive reports and monthly patient applications to best support providers in identifying opportunities for improving affordability for their patients, while at the same time supporting patient outreach, pre-visit planning, and care coordination efforts. HealthPartners has shared savings payment agreements with over 85% of its primary care providers which has increased provider engagement and sharing of appropriate risk as a partnership to lower cost for providers and patients while maintaining quality and experience.

As an example of community application, Minnesota Community Measurement as a community collaborative publicly reports Total Resource Use data by provider group in Minnesota using HealthPartners endorsed Total Resource Use measure. Through their multi-stakeholder collaborative process they collect cost data from four health plans and publicly spread the use of the measure to all provider groups in Minnesota, promoting transparency. They continue to release results annually.

While HealthPartners engages with its provider network on an ongoing basis, for external use HealthPartners has also organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download the necessary tools to run the measure, free of charge. In addition, HealthPartners has created instructions and toolkits for both SAS and non-SAS users. By creating these resources and software and putting them in the public domain it has resulted in expanded use.

U.2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1. Describe how feedback was obtained.

We continue to be engaged with provider groups on an ongoing basis throughout the year and have a partnership relationship with our provider network.

Please see specific examples of usability and quality improvement efforts shared by organizations locally and nationally under the 'Quality Improvement' section using the following link:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf

The earned shared savings providers have received over the years further emphasizes improvement. Over the years we have moved away from education of the measure and instead to how performance can be translated to operational change and targeted improvement initiatives. As technology and provider workflows evolve, we continue to enhance how we exchange information with providers to best meet their needs of incorporating information into their daily workstreams for more efficient review and use.

U.2.2.2. Summarize the feedback obtained from those being measured.

Please see specific examples of usability and quality improvement efforts shared by provider groups under the 'Quality Improvement' section using the following link:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf

U.2.2.3. Summarize the feedback obtained from other users.

Please see specific examples of usability and quality improvement efforts shared by organizations locally and nationally under the 'Quality Improvement' section using the following link:

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf

U.2.3. Describe how the feedback described in 4a2.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not

We continue to review and evaluate measure criteria based on current landscape of health care and notably as our relationship with providers has expanded to new markets. As an example, we have revised the truncation limit criteria over the years based on rising health care cost. As turnover presents within provider groups or for measure users offer education sessions and we continuously address questions and clarify detail to support measure implementation and review. We prepare user guides and maintain our public-facing website which includes several resources and technical documentation.

https://www.healthpartners.com/about/improving-healthcare/tcoc/

U.3.1. Progress on Improvement. (Not required for initial endorsement unless available.) Performance results on this measure (current and over time) should be provided in IM.1.2 and IM.1.4.

Discuss:

- **Purpose Progress (trends in performance results)**
- **Geographic area and number and percentage of accountable entities and patients included**

https://www.healthpartners.com/content/dam/brand-identity/pdfs/care/healthpartners-tcoc-usability.rtf

HealthPartners Medical Group, Park Nicollet Health Services, Essentia Health, CentraCare Health, Fairview are provider groups in Minnesota who are highlighted as engaged users of the measure and who have seen improvement in their care practices. The measure is utilized locally and nationally. Bellin Health, ThedaCare, and UnityPoint Health are provider groups in Wisconsin and Iowa who are highlighted as engaged users of the measure in their repsective regions. The details the providers have shared and the strategies they've implemented to lower cost are included in the link provided.

Since endorsement in 2012, there has been a large increase in the number of users who have adopted the Total Resource Use measure, in conjunction with the Total Cost of Care measure, resulting in improvement through greater transparency. An increase in transparency brings an awareness to the rising healthcare costs in our communities and has helped users pinpoint areas for improvement and define strategies to reduce those costs. HealthPartners provider network has been using the measure for over the past 10 years and have been managing their cost performance trend resulting in significant shared savings. Providers are fully engaged and value the opportunity to measure performance and target tangible operational change in their systems to support their patients.

HealthPartners has also organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download the necessary tools to run the measure, free of charge. In addition, HealthPartners has created instructions and toolkits for both SAS and non-SAS users. By creating these resources and software and putting them in the public domain it has resulted in expanded use.

The link above includes details of one specific example demonstrating improvement and features the Northwest Metro Alliance which serves more than 300,000 people receiving care at 9 different clinics and one hospital along with its affiliated specialists. From 2010 to 2019, the alliance achieved medical cost trends that were 6% lower than the Minnesota Twin Cities metro trend percentage. This decrease has resulted in $40 million in savings when compounding the Northwest Metro Alliance trend over 10 years compared to the Twin Cities metro.

U.3.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

We have seen great improvement demonstrated from use of the measure. There are cases in which some providers' business model is based on revenue generation and in these cases we do not see total cost or resource use improvement which also signals effectiveness of the measure's ability at indicating performance.

U.4.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

To our knowledge we are not aware of negative unintended consequences.

U.4.2. Please explain any unexpected benefits from implementation of this measure.

To our knowledge we are not aware of unexpected benefits.

Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

H.1. Relation to Other NQF-endorsed Measures

If there are related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

H.1.1. List of related or competing measures (selected from NQF-endorsed measures)

H.1.2. If related or competing measures are not NQF endorsed please indicate measure title and steward.

No related or competing measures.

H.2. Harmonization

H.2.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications completely harmonized?

H.2.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

H.3. Competing Measure(s)

H.3.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): HealthPartners

Co.2 Point of Contact: Sue, Knudson, Susan.M.Knudson@healthpartners.com, 952-883-6185-

Co.3 Measure Developer if different from Measure Steward: HealthPartners

Co.4 Point of Contact: Sue, Knudson, Susan.M.Knudson@healthpartners.com, 952-883-6185-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2003

Ad.3 Month and Year of most recent revision: 04, 2021

Ad.4 What is your frequency for review/update of this measure? Annual

Ad.5 When is the next scheduled review/update for this measure? 06, 2022

Ad.7 Disclaimers: Total Cost of Care and Total Resource Use are licensed free of charge with supporting implementation tools at the following website:

www.healthpartners.com/tcoc

Ad.8 Additional Information/Comments: HealthPartners public Total Cost of Care and Total Resource Use site: www.healthpartners.com/tcoc

For the purposes of the National Quality Forum Measure Maintenance Review for Endorsed HealthPartners Measures:

www.healthpartners.com/tcoc-documents