

National Voluntary Consensus Standards for Cost and Resource Use

FINAL REPORT

APRIL 2012



NATIONAL
QUALITY FORUM

CONTENTS

EXECUTIVE SUMMARY	2
BACKGROUND	3
NATIONAL PRIORITIES PARTNERSHIP AND THE NATIONAL QUALITY STRATEGY	4
RELATED NQF WORK	5
RESOURCE USE MEASURES IN CONTEXT	6
NQF'S CONSENSUS DEVELOPMENT PROCESS	8
Evaluating Potential Consensus Standards	8
Principles for Resource Use Measure Evaluation	8
Applying the Resource Use Measure Evaluation Criteria	9
Importance	9
Scientific Acceptability	10
Resource Use Measure Specifications	10
Approach to Disparities	19
Reliability and Validity Testing	19
Usability	22
Feasibility	22
Harmonization and Best-in-Class	23
NEXT STEPS AND FUTURE WORK	25
NOTES	26
APPENDIX A: Specifications for Endorsed Cost and Resource Use Measures 2012	27
1557: Relative Resource Use for People with Diabetes (NCQA)	28
1558: Relative Resource Use for People with Cardiovascular Conditions (NCQA)	29
1560: Relative Resource Use for People with Asthma (NCQA)	30
1561: Relative Resource Use for People with COPD (NCQA)	33
1611: ETG Based Pneumonia cost of care measure (Ingenix/OptumInsight)	36
1598: Total Resource Use Population-based PMPM Index (HealthPartners)	38
1604: Total Cost of Care Population-based PMPM Index (HealthPartners)	40
1609: ETG Based hip/knee replacement cost of care measure (Ingenix/OptumInsight)	41
APPENDIX B: Steering Committee	44
APPENDIX C: Technical Advisory Panels	45
APPENDIX D: Resource Use Measurement Terms	46

EXECUTIVE SUMMARY

As current health reform efforts focus on expanding coverage, increasing access to care, and reducing costs, it is important to understand how the system uses resources in the context of health outcomes. Combining resource use (or cost) and quality data will enable the system to better evaluate efficiency of care. Understanding resource use measurement as a building block of efficiency is a first step toward this goal. For the purposes of this project, *resource use measures* are defined as broadly applicable and comparable measures of health services counts, in terms of units or dollars applied to a population or event (e.g., diagnoses, procedures, or encounters). A resource use measure counts the frequency of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use.

This Consensus Development Process (CDP) project endorsed eight resource use (or cost) measures that will serve as building blocks for efficiency-of-care measures and signal the measure-development industry of the urgent need to develop measures of efficiency that integrate quality and resource-use domains. In applying the Resource Use Measure Evaluation Criteria for the first time, the members of the Technical Advisory Panels (TAPs) and Steering Committee encountered several overarching issues during their discussion and evaluation of these measures. Some issues varied by developer as each developer submitted measures with very distinct approaches. This report reflects the discussion of those issues as well as the measure-specific evaluation summaries for all measures reviewed during the first and second review cycles.

Eight measures were endorsed as voluntary consensus standards suitable for accountability and performance improvement:

- (1557) Relative Resource Use for People with Diabetes (NCQA)
- (1558) Relative Resource Use for People with Cardiovascular Conditions (NCQA)
- (1560) Relative Resource Use for People with Asthma (NCQA)
- (1561) Relative Resource Use for People with COPD (NCQA)
- (1611) ETG-Based Pneumonia Cost of Care (Ingenix/OptumInsight)
- (1598) Total Resource Use Population-Based PMPM Index (HealthPartners)
- (1604) Total Cost of Care Population-Based PMPM Index (HealthPartners)
- (1609) ETG-Based Hip/Knee Replacement Cost of Care (Ingenix/OptumInsight)

BACKGROUND

Per capita healthcare spending in the United States are unmatched by any country in the world.¹ This high rate of spending, however, has not resulted in better health for Americans. In fact, higher spending has not decreased mortality, increased patient satisfaction, or led to improvements in access or higher quality of care.^{2,3,4} This phenomenon of high spending with disproportionate outcomes points to a system laden with waste. The contributing factors to this alarming trend are as complex as the healthcare system itself, with physician practice patterns, regional market influences, and access to care as major drivers. Meanwhile, the United States' healthcare spending continues to increase at a rate of seven percent per year and is largely focused on treating acute and chronic illness rather than preventive care.⁵

As ongoing health reform efforts focus on expanding coverage, increasing access to care, and reducing costs, it is important to understand how resources are currently being used in the system in the context of quality, preferably related to health outcomes. Linking resource use (or cost) and quality measures will enable the system to better evaluate efficiency of care. Several provisions in the Affordable Care Act (ACA), slated to be implemented over the next five years, require using resource use data to further support efforts to move toward a value-based purchasing (VBP) payment model. One such provision requires the Secretary of Health and Human Services to develop an episode grouper that combines separate but clinically related items and services into an episode of care for an individual.⁶ Additionally, resource use data will be included on the physician compare website, as well as a physician value modifier that will be used to adjust fee-for-service (FFS) payments by combining physician performance on quality and resources use. While the ACA legislation is focused

on the Medicare population, understanding resource use measurement as a building block of efficiency, even in the context of commercial-based measures, is a first step toward meeting these goals.

For the purposes of this project, *resource use measures* are defined as broadly applicable and comparable measures of health services counts (in terms of units or dollars) that are applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure counts the frequency of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use. Current approaches for measuring resource use range from broadly focused measures, such as per capita measures, which address total healthcare spending (or resource use) per person, to those with a more narrow focus, such as measures dealing with the healthcare spending or resource use of an individual procedure (e.g., a hip replacement).

These measures will serve as building blocks for efficiency of care measures and signal to the measure development industry the urgent need to develop resource use and efficiency measures that integrate quality domains. Phase One of this work, which began in 2009, was aimed at understanding resource use measures and identifying the important attributes to consider in their evaluation. During this phase, the current National Quality Forum (NQF) Measure Evaluation Criteria used to evaluate quality measures were reviewed and refined by the Resource Use Steering Committee to address the unique aspects of resource use measures, resulting in the **NQF Resource Use Measure Evaluation Criteria**. A single steering committee was used across both phases of work, with the addition of four technical

advisory panels (TAPs) in Phase Two to assist the Committee in evaluating the measures’ clinical and methodological aspects. The consensus development process (CDP) project was divided into two review cycles, between which 14 focus areas were assigned:

Cycle 1	Cycle 2
Cardiovascular <ul style="list-style-type: none"> • Congestive heart failure (CHF) • Coronary artery disease (CAD) • Acute myocardial infarction (AMI) 	Pulmonary <ul style="list-style-type: none"> • Chronic obstructive pulmonary disease (COPD) • Asthma • Pneumonia
Stroke	Cancer <ul style="list-style-type: none"> • Breast cancer • Colorectal cancer
Diabetes	Bone/Joint <ul style="list-style-type: none"> • Hip or knee replacement • Hip or pelvic fracture • Low back pain
Non-condition specific (e.g. per capita-population)	

This report reflects the discussion and overarching issues the Committee identified while evaluating cost and resource use measures submitted to the project across two review cycles. The report is aimed at serving as foundation work for future efforts in efficiency, resource use, and cost measure evaluation.

Eight measures were endorsed as voluntary consensus standards suitable for accountability and performance improvement:

- (1557) Relative Resource Use for People with Diabetes (RDI) (NCQA)
- (1558) Relative Resource Use for People with Cardiovascular Conditions (NCQA)
- (1560) Relative Resource Use for People with Asthma (NCQA)
- (1561) Relative Resource Use for People with COPD (NCQA)
- (1611) ETG-Based Pneumonia Cost of Care (Ingenix/OptumInsight)
- (1598) Total Resource Use Population-Based PMPM Index (HealthPartners)
- (1604) Total Cost of Care Population-Based PMPM Index (HealthPartners)
- (1609) ETG-Based Hip/Knee Replacement Cost of Care (Ingenix/OptumInsight)

NATIONAL PRIORITIES PARTNERSHIP AND THE NATIONAL QUALITY STRATEGY

The **National Priorities Partnership** (NPP), a multi-stakeholder collaborative of 48 organizations convened by NQF, plays a key role in identifying strategies for achieving national goals for quality healthcare and facilitating coordinated, multi-stakeholder action. The Department of Health and Human Services has asked the partnership for its collective, multi-stakeholder input on the **National Quality Strategy** (NQS) framework, which includes three inextricably linked domains—better care, affordable care, and healthy people/healthy communities—around which priorities,

goals, measures, and strategic opportunities for improvement are to be identified or refined.

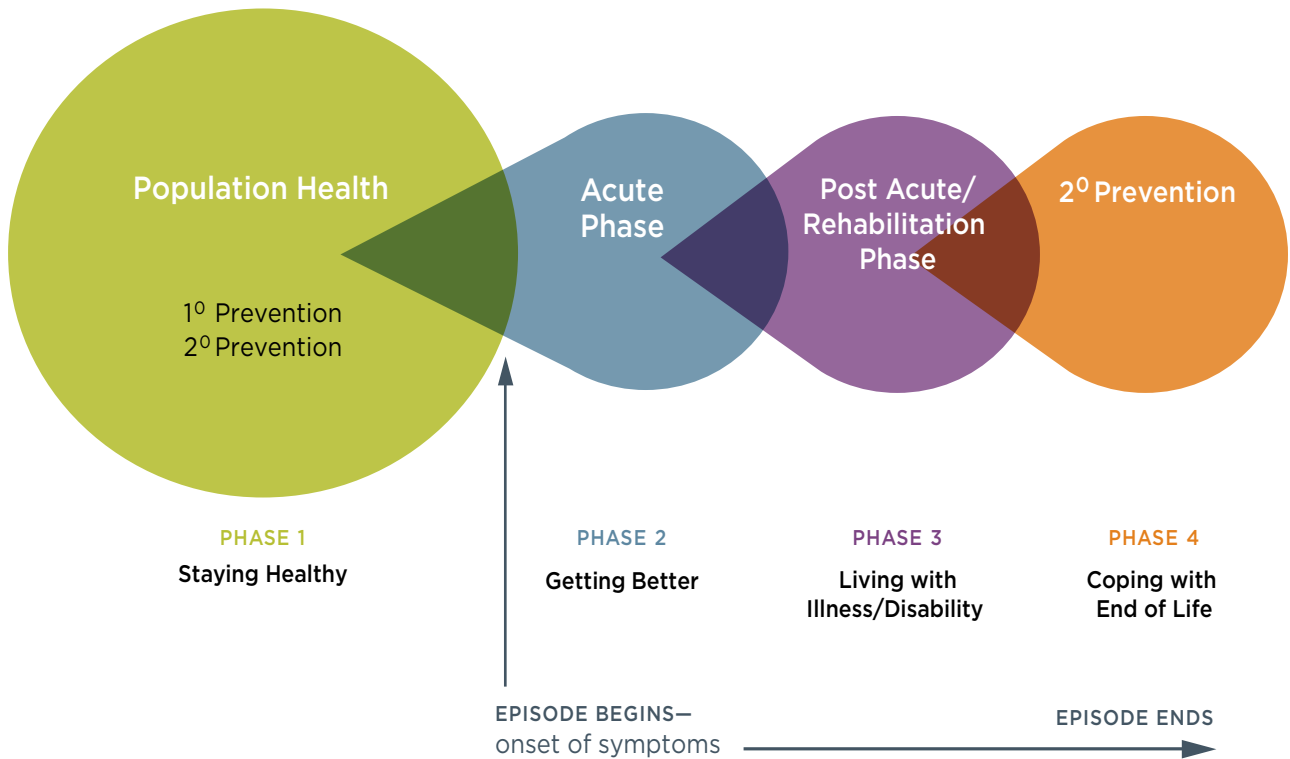
When the NQS was announced in March 2011, one of the priorities it identified was making quality care more affordable. The resource use measure endorsement process is an important step toward measuring affordable care by evaluating resource use and cost measures. These measures can identify opportunities to reduce the rate of growth in healthcare spending, and when paired with quality measures, can help evaluate the efficiency of the healthcare system.

RELATED NQF WORK

This project is NQF’s first effort focused on evaluating and endorsing cost and resource use measures. In 2009, NQF completed a measurement framework for evaluating efficiency across patient-focused episodes of care (*Figure 1*). This report, **NQF Measurement Framework: Evaluating Efficiency across Patient-Focused**

Episodes of Care, presents the NQF-endorsed® measurement framework for assessing efficiency, and ultimately value, associated with the care over the course of an episode of illness and sets forth a vision to guide ongoing and future efforts.

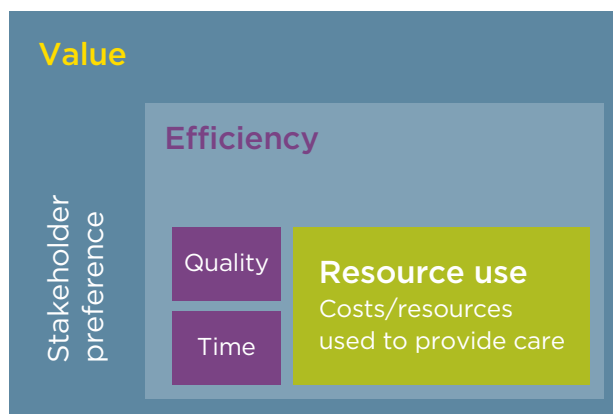
FIGURE 1. A GENERIC EPISODE OF CARE MODEL



RESOURCE USE MEASURES IN CONTEXT

This CDP seeks to endorse resource use (or cost) measures as building blocks toward measuring efficiency of care as illustrated in *Figure 2*. Efficiency can be defined broadly as the resource use (or cost) associated with a specific level of performance with respect to the other five Institute of Medicine (IOM) aims of quality: safety, timeliness, effectiveness, equity, and patient-centeredness.⁷ Time is sometimes used to define efficiency when determining efficiency of throughput processes or applying time-driven activity based costing methods. Resource use measures can also be used to assess value by integrating preference-weighted assessments of the quality and cost performance of a specified stakeholder, such as an individual patient, consumer organization, payer, provider, government, or society.⁸

FIGURE 2. RESOURCE USE AS A BUILDING BLOCK OF EFFICIENCY AND VALUE



As a building block in understanding efficiency and value, NQF supports using and reporting of resource use measures in the context of quality performance, preferably outcome measures. Using resource use measures independent of quality measures does not provide an accurate assessment of efficiency or value and may lead

to adverse unintended consequences in the healthcare system.

Resource use measures used to assess efficiency and value should be important to measure, have scientifically acceptable properties, and be usable and feasible. Resource use measures under evaluation in this process should independently meet these endorsement standards. Future efforts will need to evaluate how resource use measures can be linked with appropriate quality measures to assess the healthcare system's efficiency. These efforts should consider quality and resource measure alignment of the underlying population, exclusions, and risk adjustment, among other measure properties.

Given the diverse perspectives on cost and resource use measurement in healthcare, it is important to articulate, in the context of this project and the measures submitted, the terminology, purpose, and perspectives these measures represented. Recognizing this is NQF's first project in the resource use measurement arena, there is a clear gap in the NQF portfolio for these types of measures. NQF also recognizes that while the measure submission process is open to any entity wishing to submit measures for evaluation, the measures submitted and evaluated in this process are not representative of all approaches to measuring healthcare costs and resources that exist in the market today. This report is a reflection of the evaluation process of the measurement approaches submitted to this project for review.

Each of the measurement approaches submitted for review calculates the use of various resources using administrative claims data, then categorizes them by type of resource (e.g., pharmacy, durable medical equipment, evaluation and management (E&M) visits) and applies a costing methodology

(either actual prices paid or standardized prices). When developers further apply a dollar value to utilization counts, the dollar value serves as a weight for each resource. Due to the limitations in the data types available for measuring resource use in healthcare, administrative claims data are the primary source of this information for the measures submitted to this project.

Also important to understand in the context of this report is the way in which the terms “cost,” “resource use,” and “prices” are used. The term “cost” can represent very different constructs to various stakeholders. In the context of this report, cost (or cost-of-care measures) reflects the actual price paid by health plans for health plan member utilization; resource use (or resource use measures) further applies standardized prices to utilization counts. By many accounts, prices charged by providers is not an accurate measure of utilization as prices charged vary and can be a reflection of the negotiating position of health

plans vis-à-vis providers in a given market⁹. Actual price paid is generally a reflection of the cost the health plan incurs to cover the claims submitted for its members; some measures also include a member (consumer) cost based on member co-pays, coinsurance, and deductibles. For a provider, (e.g., a physician or nurse practitioner) a cost-of-care measure would reflect the payment the provider received from the health plan for care provided. For a purchaser, a resource use measure can be used to assess the utilization of healthcare services across health plans, while a cost-of-care measure can be used to assess how well a health plan is managing charges and utilization of providers within the health plan’s network. Given the other types of costs attributed to healthcare, it is important to note that these measures do not capture or represent production costs (fixed or variable), administrative costs, government funding to support healthcare delivery, or societal costs (e.g., lost wages, sick days).

NQF'S CONSENSUS DEVELOPMENT PROCESS

NQF's National Voluntary Consensus Standards for Cost and Resource Use project seeks to endorse resource use and cost measures for performance improvement and accountability in the context of quality measures.

Evaluating Potential Consensus Standards

Candidate consensus standards were solicited through a Call for Measures on January 31, 2011. Across the two review cycles, 36 measures were submitted and evaluated for suitability as voluntary consensus standards for accountability; 21 of these were withdrawn by the developer for further refinement and testing. The measures were evaluated using **NQF Resource Use Measure Evaluation Criteria**. Four condition-focused TAPs for pulmonary, cardiovascular and diabetes, bone and joint, and cancer conditions rated each candidate consensus standard according to the subcriteria. The TAPs then identified strengths and weaknesses to assist the committee in making recommendations. The 23-member, multi-stakeholder committee evaluated the subcriteria of the non-condition specific measures, provided final evaluations of the four main criteria—importance to measure and report, scientific acceptability of the measure properties, usability, and feasibility—and made endorsement recommendations for all measures. Measure developers were available during TAP and committee discussions to respond to questions and clarify any issues or concerns.

Principles for Resource Use Measure Evaluation

In phase one of this project, the committee defined resource use measures and their constructs to better understand how to evaluate these measures. Also within this phase of work, the committee developed the following principles

to frame the evaluation of resource use measures for endorsement:

1. Efficiency is one of the Institute of Medicine (IOM) five quality aims and is a function of resource use and health outcomes: *Efficiency = $fx(\text{resource use, health outcomes})$*
2. Resource use measures are the amount of resources used per population, episode, or procedure.
3. Resource use measures are an important building block for measures of efficiency of care; future measurement efforts should integrate and explicitly incorporate measures of quality, health outcomes, or appropriateness.
4. The justification for and intended purpose of resource use measures is to examine, understand, and ultimately reduce unnecessary costs in care.
5. There is a continuum of resource use measures (i.e., per capita to per procedure); all types under consideration for endorsement must meet NQF evaluation criteria for such measures.
6. The resource use measure specification and calculation must be explicitly stated and transparent so the approach can be deconstructed and implemented in a standard manner.
7. Comprehensive measures are preferable, even if combining multiple service categories into one resource use estimate increases complexity; using methodologically sound methods is of paramount importance.
8. The final resource use measure result should be clear and understandable for all stakeholders to interpret.
9. Methods for combining the component scores influence the interpretation of the measure results and must be justified (e.g., averaging across all component scores may obscure low or high scores of individual components).
10. While resource use measure developers may have fundamental differences in approach, these

principles should apply across all types and approaches.

11. NQF considers transparency key to ensuring the intended audiences understand the results and can use them for decision making. Resource use measures are often highly complex, with lengthy algorithm decision trees that can make clarity difficult, particularly when some components may be only partially transparent to the user.

Applying the Resource Use Measure Evaluation Criteria

With a working definition of resource use measures and guiding principles in place, the committee completed a detailed review of the standard NQF Measure Evaluation Criteria. This review resulted in the [NQF Resource Use Measure Evaluation Criteria](#), based on the same four major criteria used to evaluate quality measures—importance, scientific acceptability, usability, and feasibility—with targeted changes to the subcriteria to address the unique attributes of resource use measures.

In applying the Resource Use Measure Evaluation Criteria for the first time, the TAPs and committee encountered several overarching issues during their discussions and evaluations of the measures. Some issues varied by developer, as each developer submitted measures with very distinct approaches. The committee factored these issues into its ratings and recommendations for multiple measures, recognizing the need to balance the quantity and specificity of information required to adequately evaluate the measure and the burden on the developer to provide this information.

At the conclusion of the measure review process, there was a great deal of learning to reflect upon and a unique opportunity to provide recommendations for future efforts in this area. While resource use measurement has been used in the commercial sector for many years, the emerging interest in using these measures for public reporting and payment initiatives suggested an increased need for a transparent

peer review of candidate resource use measures. Based on their experience reviewing the measures submitted to this project, the committee was asked to provide guidance to the field for future efforts to develop and evaluate resource use measures. Through this exercise, the committee offered recommendations related to clarifying the submission process, improving data quality, measuring resources in the Medicare population, and linking quality and resource use measures. Additionally, the committee raised several issues around risk adjustment, reliability and validity testing of resource use measures that aligned with the guidance laid out in the NQF testing task force report.

A summary of the committee's considerations during the measure evaluation process and related recommendations are discussed in the context of each of the four major criteria. Summaries of the discussion for each measure can be found in the resource use technical report on the NQF website.

Importance

The importance criterion for resource use measures, like that for quality measures, is aimed at determining the extent to which the measure's focus (e.g., hip fractures, coronary artery disease) is important to measure and report. For resource use measures, the developers were asked to demonstrate high impact by showing there is variation and opportunities for improvement in the delivery of care for the identified condition. The TAP concluded that the measures submitted were broad and inclusive of high-impact conditions. Additional subcriteria were tailored specifically for resource use measures. These subcriteria included an evaluation of whether the intent of the measure had been clearly described and whether the resource use service categories selected to measure costs accurately reflected the intent and focus of the measure. Most measure submissions were found to be important due to the gaps in measurement in this area and the need to understand resource use and costs broadly

and in the condition-specific areas. However, in one instance, the Bone/Joint TAP reviewed a hip fracture measure that was specified and tested in a population with an age distribution outside of the age range in which the condition was most prevalent. The TAP agreed this approach calls into question the importance (and in fact the validity) of the measure as it has been tested and used to measure costs in a population where this condition is not high impact and has limited clinical relevance.

Scientific Acceptability

Similar to quality measures, evaluating the scientific acceptability of resource use measures includes reviewing the measure's specifications, reliability and validity testing, and approach to addressing disparities. The completeness, repeatability of the specifications, and the adequacy of the reliability testing methodology and results are evaluated within the reliability criterion. Applying the validity criteria, the committee was asked to determine whether the specifications reflected the intent of the measure and address those areas where there was variation. The validity criterion also includes an assessment of the adequacy of validity testing, exclusions, risk-adjustment, and the identification of meaningful differences.

Resource Use Measure Specifications

Within the foundational phase of this work, the committee determined that the resource use measure specifications submitted to this project should be delineated by five modules: 1) data protocol, 2) measure clinical logic, 3) measure construction logic, 4) adjustments for comparability, and 5) measure reporting. To allow for user flexibility, the developers were permitted to submit measurement steps in the data protocol and reporting modules as specifications or guidelines, or to not submit instructions at all. Specifications are inherent measure characteristics that must be fully implemented to obtain valid measure results. Guidelines, on the other hand,

are suggested approaches from the developer on possible ways to implement these steps.

Reflecting on this approach of allowing specifications and guidelines in the submission, the committee identified some challenges and recommendations for future efforts. In an effort to minimize the potential for confusion in the submission and evaluation processes, the committee identified areas within the resource use measure specification modules that should be clarified to ensure the developer has full understanding of the information required for the measure to be considered. The committee recognized that in order to improve the clarity of the measure submissions, there should also be attention paid to how new submission requirements will affect the burden on the developer to submit measures for consideration. While there are some areas of the submission that will need additional information and more clarity, there may be components of the measure submissions from that may not be required.

While guidelines in measure components may be acceptable to promote measurement for comparison across entities nationally, instructions relevant to the data protocol, measure clinical logic, construction logic and adjustment for comparability, should be standardized in the form of specifications.

Evaluation of resource use measure specifications proved to be the most intensive effort in the review process. The issues identified within each of the specification modules have been outlined below.

DATA PROTOCOL

The data protocol module allows developers to submit instructions and analytic steps for

cleaning or aggregating relevant data necessary to implement the specifications and produce valid results. Measure developers submitted the following data protocol information: data preparation (e.g., types of data required, continuous enrollment requirements), data inclusion criteria (e.g., number of months of claims data needed), data exclusion criteria (e.g., instructions for rejected, \$0, or high-dollar claims), and considerations for missing data (e.g., instructions for imputation). Recognizing that not all developers create specifications around these steps, the committee agreed during phase one of the project, these items could be submitted as specifications or guidelines, or not submitted at all. However, during the evaluation process, allowing for flexibility in this module led to some discomfort for the experts, specifically related to instructions for handling missing data. Ensuring that the data used to run the resource measures are complete and representative is a critical first step to generating valid measure results. Allowing flexibility in these steps could allow for errors and inconsistent implementation of the measure. This unanticipated concern led the committee to recommend going forward that the instructions within the data protocol module be submitted only as specifications.

Carve-out arrangements

A major concern of the committee throughout the evaluation process was data limitations and the implications of incomplete data, specifically, the

Comparisons of entities with and without carved-out data is inappropriate.

impact of carve-out arrangements on accurately capturing resources used. Accountable entities may outsource services through pharmacy benefit managers (PBMs) or behavioral/mental health carve-outs, which may result in incomplete or

missing pharmacy or behavioral/mental health data. These entities can outsource administration of outpatient prescription drug benefits to PBMs. Carve-out arrangements allow accountable entities to separate behavioral/mental health insurance benefits by contracting with a third party to manage care or the insurance risk for patients requiring these services.¹⁰ The committee agreed that total resource use for entities that do not receive member claim information from carve-out pharmacy and behavioral/mental health services may not be comparable to resource use for those that do not outsource these services. In this instance, interpreting the overall costs for a patient across health plans with and without carve-out arrangements would be misleading.

However, entities without member claims data from their carve-out arrangements can be flagged for comparison with entities with similar missing benefit information. Because resource use

Measure scores calculated and reported using data with carve-outs should be labeled as such.

measures allow claims to be assigned to resource use categories (e.g., laboratory and imaging), these categories can be used to compare costs across entities, even when outsourcing arrangements are present. For example, comparing laboratory costs or imaging costs across entities within a total per-capita resource use measure would be informative even when pharmacy data are not available.

Further, when developing resource use measures, careful consideration should be given to whether the importance of measuring resources/costs in an area outweigh the limitations of the data. For some conditions, the lack of robust data could distort the measure output. For example, to measure the resources for asthma patients where greater than 40 percent of the resource use may be pharmacy related, data sets without pharmacy

data are inherently misleading in providing useful insight into the cost of asthma care.¹¹ For acute or procedural episodes (e.g., hip replacement)

If a measure is intending to measure a clinical condition that has a predominant portion of its costs in pharmacy claims, consider whether costs should be measured at all in the absence of these data. The developer is responsible for determining whether the lack of these data invalidates the measure score for comparisons.

where the care is more standardized (e.g., pre- and post-surgical antibiotics) pharmacy and mental health data do not account for a major portion of the resource use and thus administrative data, and carve-out concerns may not have a tremendous impact on the measure results. In future efforts, the committee suggested that the resource use submission form explicitly request that the developer address how carve-outs are identified and reported.

Use of administrative data

All of the measures submitted use administrative claims as the primary data source. Administrative claims offer the benefit of reduced administrative burden for providers and measure implementers in collecting and reporting data elements. However, variation in coding practices has the potential to affect the reliability and validity of any measure that relies on administrative and claims data alone, including resource use measures. This may be particularly true for entities providing care under capitated financial arrangements that may capture fewer diagnostic and procedural codes per record than those operating under traditional fee-for-service arrangements.

In an effort to address some of the underlying global issues affecting the use of administrative

claims data for the purposes of measuring resource use, the committee identified several areas in which healthcare stakeholders might engage and support additional efforts to improve the ability of resource use measures to capture all resource use fully. The use of administrative claims data presents certain limitations for measuring resource use performance, limitations that are present in quality performance measurement as well. Primarily the reliance of resource use measures on administrative claims data to count resources, or dollars spent, captures only the output on behalf of the provider—not the costs to the patient, nor the costs or resources for which there are no administrative codes. Recognizing this as a limitation of the data available to measure these types of resources, the committee recommended that future efforts in resource use measurement focus not only on the costs to the provider, but to the user as well, through identifying those resources that are important to measure and determining how to capture this data. The committee recognized that while administrative data are the primary data source used for measuring resources at this time, there is opportunity to integrate the data gathered through electronic health records (EHRs) and other clinical data to measure resource use.

Since resource use measurement is a priority for many stakeholders, efforts should be made to ensure the necessary data are available for accurate measurement. However, there are significant challenges to determining where the

Data sets used to measure resources should be as comprehensive as possible. Efforts to obtain clinical and carved-out data (e.g., pharmacy, behavioral health) should be made to ensure the data set used to calculate resource use is robust, complete, and representative.

responsibility lies to ensure data are complete and the ways in which important, but sensitive information is shared. For a number of measures submitted for evaluation in this project, the instructions within the data protocol module suggest that the measure implementer is responsible for ensuring data are complete and representative. The committee acknowledged however that measure implementers often do not have the resources or technical expertise to audit data before use. Future efforts should explore a potential role for large data aggregators to identify thresholds and set standards for data quality.

CLINICAL LOGIC

Evaluation of the measure clinical logic included steps to identify the condition or event of interest and any clustering of diagnoses or procedures. This evaluation included examining the clinical topic area and determining whether or not the measure accounts for comorbid conditions, disease interactions, clinical hierarchies, clinical severity levels, and concurrency of clinical events.

The complexity of the submitted measure specifications made evaluating the measure's clinical logic challenging. For example, measure developers designed various methodologies to assign patients to a severity level; however, due to complex algorithms, specific details and code lists used to determine the assignment of patients to severity categories were difficult to interpret.

Exclusions were a focus during evaluation of the resource use measure's clinical logic. Although the creation of homogenous populations enables comparability, measure developers should ensure that measure exclusions do not allow for complications from poor care to drive patients out of the episode, thus rewarding entities that provide inadequate care. For example, a biased measure score may be created by excluding patients with acute myocardial infarction (AMI) who are discharged from a skilled nursing facility or excluding patients who are not discharged alive.

Finally, resource use measures that seek to create more homogenous patient populations often are limited by the ability of administrative claims data to assess patient health status and severity accurately. For example, measures submitted were unable to differentiate between community-acquired and healthcare-acquired pneumonia. Measures submitted also were unable to identify staging information to assess the severity of a cancer diagnosis.

CONSTRUCTION LOGIC

The measure construction logic evaluation included a review of the steps used to cluster, group, or assign claims beyond those associated with the measure's clinical logic and an assessment of how the various components of the measure (episode logic, clinical logic, risk adjustment) work together. Measures were evaluated to determine if the temporal parameters including trigger and termination rules are appropriate for the clinical logic specified within the measure. For example, the committee evaluated the post-hospitalization period in an episode of AMI to ensure it was appropriate for the measure's intent, level of analysis, attribution approach, and statistical properties.

The committee evaluated the validity of the measures by examining the interaction of the measure components including the specified level of analysis and the risk adjustment approach. There is a need for nationally endorsed measures at the individual clinician level of measurement and the experts encourage development of measures at this level. However, the committee expected developers to demonstrate statistical differences at sample sizes that would be observed in the level of analysis specified. Further, attribution of the measure to the individual or group practice level was discussed at length, focusing on the appropriateness and generalizability. While sample size and attribution could be submitted as guidelines, the committee agreed these testing results contribute to the measure's scientific acceptability.

Measures submitted as a part of an episode grouper were challenging to evaluate because the assignment of claims into the episode, comorbidities and interactions, clinical hierarchies, and the handling of concurrent clinical events included lengthy algorithm decision trees that were at times unclear and only partially transparent to the reviewers. Measures submitted to this project were evaluated as standalone measures of resource use. However, the construction logic within episode grouper-based approaches includes claim assignment decisions, or tie-breaker logic, which was not clearly explained in the evaluation of single resource use measures. Tie-breaker logic is a mechanism to determine how a claim or record is assigned to an episode if it is eligible for assignment to multiple episodes. For example, if a patient fills a prescription that could be mapped to multiple open episodes, tie-breaker logic could be used to determine to which episode this cost would be assigned. The committee expected developers to provide a clear and transparent explanation of this tie-breaker logic, how claims would be assigned to episodes, and how various open episodes interact with each other. While resource use measures are complex, developers have a responsibility to provide an explanation of the construction logic within the grouper.

Multiple co-occurring episodes

The challenge of evaluating condition-focused episodes as part of a grouper and in the context of real-world patients who often have multiple co-occurring conditions, highlighted the challenges of applying this approach to measure costs in the Medicare population where individuals with multiple chronic conditions account for the majority. More than half of all beneficiaries were treated for five or more conditions, accounting for three-fourths of total Medicare spending.¹² In 2002, more than 92.2 percent of all Medicare healthcare spending was by beneficiaries with three or more conditions during the measurement year.¹³ Resource use measures for the Medicare

population will have to consider how to count costs for treatment of multiple co-occurring conditions, and how to integrate data and attribute costs from multiple sites where beneficiaries seek care and account for resource use at the end of life. There will be an urgent need for measures specified for the Medicare population for use in bundled payment demonstrations, physician feedback reporting programs, and value-based purchasing programs.

Episode approaches attempting to assign claims to specific episodes should create a transparent hierarchy with rules to assess resource use in the Medicare population accurately. One approach suggested in the committee's discussion would allow flexibility in the assignment of individual claims to a single episode or to multiple open episodes. This patient-centered approach could allow an individual office visit for evaluation and management to be assigned to multiple episodes. Cost estimates should be based on the time and attention a provider should be reasonably expected to deliver on a patient's multiple co-occurring conditions beyond the acute disease and its immediate complications for which the patient sought care.

Efforts to develop resource use measures for the Medicare population should consider the NQF consensus measure framework for assessing the efficiency of care for individuals with multiple chronic conditions (MCCs). **MCC framework** guiding principles include promoting shared

A patient-centered approach should be used to describe the interaction of conditions (and episodes) in the development of resource use measures for the Medicare population.

accountability with members of the healthcare system, a multi-dimensional measure approach that incorporates various types of measures, a

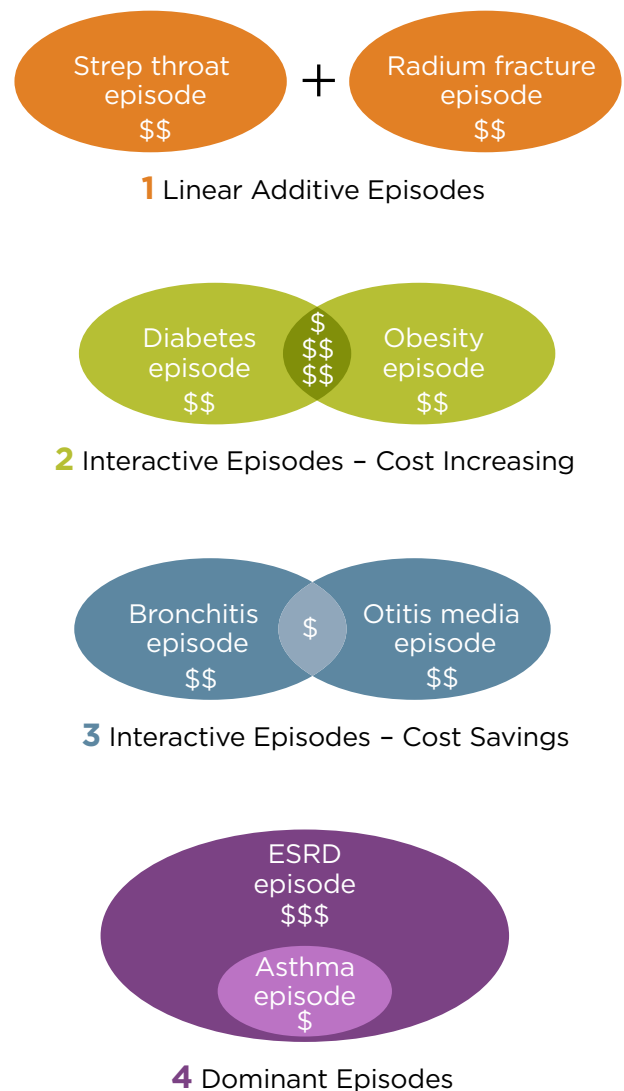
focus on shared decision making in concordance with a patient's preferences, and prioritization of measures across time that are most relevant to achieving desired outcomes as determined by the care plan.

The committee recognized that the contribution of individual conditions to the total cost of managing a beneficiary may vary, depending on other conditions present in each beneficiary. The following classification system of four types of overlapping episodes helps to illustrate this (*Figure 3*): 1) linear additive episodes, 2) interactive episodes-cost increasing, 3) interactive episodes-cost savings, and 4) dominant episodes.¹⁴ Linear additive episodes occur when the patterns of illness are not overlapping, and episodes can be considered independent of one another. For example, a fracture of the radius and strep throat would be considered independent of one another.¹⁵ Interactive episodes can increase costs when there are two or more conditions in which the presence of multiple conditions increases the level of resources required to treat all of the conditions. An example would include the treatment of diabetes in the presence of obesity.¹⁶ In this situation, the cost of the combined condition is more than the sum of the individual parts. Interactive episodes can also be cost saving since the cost of treating overlapping conditions is not likely to require significantly different resources (e.g., the treatment of otitis media and bronchitis).¹⁷ Finally, in dominant episodes, the dominant condition becomes the principle focus of care in the presence of a mild disease (e.g., the treatment of end-stage renal disease in the presence of mild asthma). These methods for overlapping episodes should be considered in developing approaches for assessing resource use in the Medicare population.

When developing measures to assess the resource use for a chronic condition, the nature of the interaction and resource use for an acute complication for that condition should be considered. The committee considered the

example of misinterpreting lower CAD resource use as better performance when, in fact, a per-capita assessment may demonstrate higher resource use. The higher resource use may be derived from higher rates of AMI in the measured population due to poor CAD management.

FIGURE 3. CLASSIFYING OVERLAPPING EPISODES, ADAPTED FROM HORN BROOK (1985)



With the advent of accountable care organizations (ACO's), additional efforts are needed to propose alternative attribution approaches to encourage team-based care along the patient episode of care. Resource use measures developed for the Medicare population should also consider that

beneficiaries often seek care from multiple sites. The typical Medicare beneficiary sees two primary care physicians and five specialists working in four different practices.¹⁸ The committee discussed how current attribution models assign treatment of the patient to an individual provider based on the number of visits or the highest proportion of costs. However, in a patient-centered model, all providers who treat the patient should have responsibility for the care delivered.

Another important consideration for developing episodes of care for the Medicare population involves accounting for utilization for end-of-life (EOL) care. Simply including EOL patients in estimates of episode-based resource use has the potential to introduce inappropriate incentives. Resource use measures that include EOL patients should be reported with balancing mortality measures to ensure that providers are not inadvertently reported as providing more efficient care when they have higher rates of mortality. On the other hand, with resource use during the last year of life accounting for more than a quarter of Medicare payments,¹⁹ EOL patients should not be excluded from the analysis of resource use. Future evaluation of resource measures for the Medicare population should consider how measure developers handle EOL patients in profiling providers.

ADJUSTMENTS FOR COMPARABILITY

A measure's result can be influenced by confounding external factors that can affect the measure score. Measure developers submitted steps for adjusting measures to increase comparability. These adjustments include risk adjustment, stratification approach, and the costing method used within the measure.

Risk adjustment

Risk-adjustment methodologies varied considerably across measure developers. A combination of complexity and a varying degree of transparency of the risk-adjustment approach made evaluating the methods

challenging. The experts agreed that the details on the performance of risk models were vital to determining the model's adequacy—specifically, how the presence of certain claims drives categorization into risk categories and the goodness of fit of the risk model. Of the various methodologies reviewed, none was considered to be superior. A **Society of Actuaries (SOA) report** was shared with the committee comparing various risk-adjustment methodologies (e.g., Hierarchical Clinical Categories (HCC), Adjusted Clinical Groups (ACG), Episode-risk-group (ERG)) was informative. However, more research and guidance on the appropriateness of the models for specific applications are needed, as the committee deemed this report to be an inadequate analysis of the risk-adjustment models for the purposes of this project. Guidance presented in the SOA report was insufficient in assisting the committee's assessment of risk-adjustment model performance across various datasets and across various homogenous populations (including Medicaid or Medicare). It was also insufficient in assisting in the assessment of, or the credibility of risk-adjustment models across various population sizes.

After reviewing various risk-adjustment approaches presented in the measures submitted to this project, the committee agreed measure developers need to demonstrate that the specified risk models are appropriate for the target population. For instance, if the hierarchical condition category (HCC) model is used to measure a commercial population, developers

The risk adjustment model applied to the measure should be specific to the intended population (e.g. commercial, Medicare, Medicaid).

should demonstrate that it is appropriate for use outside of a Medicare population. The committee agreed that risk models have unique weights for

comorbidities and may not include all relevant conditions (e.g., pregnancy) when the risk-adjustment model is used outside of the population in which it is calibrated. Measure developers have the burden of demonstrating appropriateness through R-squared values and through a detailed clinical and statistical explanation of how variables were added to the risk model. Additional research is needed in this area to explore how various risk-adjustment approaches change the relative ranking of providers in terms of resource use and how the use of clinically enhanced administrative data may impact measure scores and the selection of factors added to the risk-adjustment models.

The committee agreed that measure developers need to demonstrate that variables included in the risk-adjustment model are not simply selected based on their statistical explanatory power, but rather, risk factors are well documented in clinical evidence. When variables are chosen for inclusion in the risk-adjustment model, developers are responsible for demonstrating a relationship to the outcome of the measure (i.e., resource use). Additional detail through a sensitivity analysis including various risk-adjustment variables can be provided in future evaluations to demonstrate the effect of variables included in the final risk-adjustment approach.

Stratification is a method that can be employed to create homogenous risk populations. However, similar to the concern that exclusions may remove patients out of an episode inappropriately, measure developers need to ensure that the risk stratification approach does not allow for complications from poor care to drive patients into a higher risk stratum, thus rewarding entities who provide inadequate care. For example, for patients with coronary artery disease (CAD), creating risk strata based on subsequent revascularization has this potential for adverse consequences.

Costing methodologies

The adjustments for comparability module also include the costing approach (i.e. how utilization

counts should be monetized). The appropriateness of costing approaches was a major issue of discussion throughout the evaluation process. For the measures submitted, the costing approaches were either specified for the actual prices paid (i.e., cost of care measures) or for standardized prices (i.e., resource use measure). Standardized pricing allows users to compare the use and intensity of health services while holding actual paid amounts constant. Resource use measures that apply standardized prices allow for comparison of resource use units across regions and markets, while actual prices allow for comparison of prices paid.

While the committee agreed that both costing approaches could be appropriate for different applications, they agreed that a measure should only include one costing approach. One costing approach allows for standardized implementation and ensures consistent and accurate comparisons of measure results. While the combination of these approaches in a single measure is typical for use in the commercial sector, for use as a national consensus standard, measure results should unambiguously reflect differences in performance for an accountable entity, not differences in the type of data that an entity chooses to submit (actual prices or standardized prices). Developers who submitted a single measure with an option for the user to determine which costing method to apply were asked to either split the submission into two separate measures or select one of the approaches to apply to a single measure submission.

Subsequent committee discussions on applying an actual price approach for national comparisons at an individual provider level identified additional concerns. Specifically, the committee noted the potential for misinterpreting physician resource use in national reporting since this pricing approach includes environmental factors (e.g., local facility and labor costs) that may be outside of an individual clinician's control. The committee agreed that when actual prices paid is reported, utilization counts should be reported as well.

Concern over the use of actual prices also was considered in the measure's usability.

There was agreement that actual prices paid by health plans to providers is important to measure and report. For example, regional comparisons at the individual provider level where environmental factors may not be as prominent, or nationally at higher levels of measurement (e.g. health plan level). Regional comparisons of pricing variation using measures of actual prices paid allow stakeholders to monitor for an increase in the price for health care services.

Finally, measures submitted to this project spanned various levels of measurement analysis, from regional, to health plan, to individual provider. Measures specified at a higher level of measurement (i.e., health plan or regional) allowed for a comprehensive view of health service resource use by measuring all costs for a person across settings and providers. While the committee encouraged measurement at the individual clinician and group practice level, measures submitted to this project had difficulty demonstrating reliability and validity at this level.

REPORTING

The reporting module includes steps for attribution, peer grouping, defining outliers and thresholds, sample size requirements, and benchmarking. These reporting steps could be submitted as measure specifications or guidelines, or could be left to the user's discretion. Specifications limit user options and flexibility and must be strictly adhered to, whereas guidelines are well thought-out guidance to users, allowing for user flexibility. While this type of information is not collected for quality measures, the committee determined in phase one that for resource use measures, this type of information was uniquely important to the implementation of a resource use measure. However, similar to the data protocol module, reviewing components of the measure submitted as guidelines posed some challenges. Within the reporting module, the committee was very concerned with how the reviewed models

might be applied, even as guidelines.

While sample size considerations could be submitted as guidelines or specifications in the reporting module, the committee found that sample size was also relevant to the discussion of other modules and reliability and validity testing. To evaluate the number of patients required for a measure to demonstrate meaningful and statistically significant differences, the committee encouraged measure developers to provide simulations and sensitivity analyses during the evaluation. Across the various measurement approaches, outliers were handled at both the episode and the claim level. During data preparation, high outlier claims were generally subject to a statistical technique used to limit the effect of extreme values and the effect of spurious outliers, known as *winsorization*.²⁰ Low cost claims were either winsorized or, more typically, were removed from measure analysis. Winsorization often sets outliers to a percentile of data; for example, all outliers above the 95th percentile are set to the value at the 95th percentile. Developers who chose to remove low-cost episodes indicated they took this approach because these episodes were likely to be incomplete and thus have the potential to skew the results. The committee requested additional details from the developers on the effect of the winsorization and exclusion at the claim and episode level on the measure score. The experts noted that detailed listing and analysis of high-cost outliers could be useful for targeted improvement activities.

As part of the reporting module, the attribution approach could also be submitted as measure guidelines or specifications or left to the user to define. Each developer submitted their measures with the attribution approach(es) as guidelines. The attribution approach is distinct from the level of analysis in that the level of analysis is the unit in which the measure has been tested and specified, while the attribution approach determines how the costs or resources are assigned to a provider, group of providers, health plan, or region.

Further, users need flexibility in the approaches to accommodate specific applications and the opportunity to consider input from the attributable entities. In reviewing several of the attribution guidelines, the committee did note that proper consideration should be given to how the timing of patient encounters affects the attribution rules and potential for unfair assignment of costs to clinicians. For example, if the attribution approach assigns a patient to the primary care provider (PCP) based on one evaluation and management (E/M) visit, the approach should not assign all of the previous hospitalization costs during the measurement year before the patient's first visit to this PCP. Lack of consideration for these types of factors may create unintended consequences for patients seeking primary care after high-cost hospitalizations or procedures.

With no accepted gold standard for attribution, and a lack of widespread agreement on any of the attribution approaches reviewed, the committee recognized that there must be some attribution as healthcare systems may require varied approaches for their unique market. This highlights the need for more discussion on how, if at all, attribution approaches should be evaluated in this process where the goal is to endorse standardized approaches to measurement.

Approach to Disparities

Identifying and measuring disparities in care delivery is critically important to understanding variations in cost and improving quality. Gender and age were the most common factors accounted for in the stratification for disparities in the measures reviewed. The lack of information on race and ethnicity in commercial administrative data limited the ability of the resource use measures under evaluation to reflect disparities accurately in the results. Additional efforts should be pursued to capture this information more systematically. As such, the committee was unable to assess the measures' ability to identify disparities based on underlying limitations in the data. Measures were evaluated based on their

ability to stratify if the underlying data included information on race and ethnicity.

Reliability and Validity Testing

The next component to evaluating a measure's scientific acceptability is determining whether the measure testing approach and results demonstrate that the measure is reliable and valid. Reliability and validity testing is included in the NQF evaluation criteria, and NQF allows flexibility in the specific methods used in testing to allow measure developer flexibility. The committee evaluated: 1) the scope of testing, 2) what tests of reliability and validity could be performed, and 3) how to weigh the results of this testing. The steering committee interpreted testing results within the unique context of the specific measure under review

Reliability testing should demonstrate that the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period, or that the measure score is precise. Validity testing must demonstrate that the measure data elements are correct or that the measure score correctly reflects the cost of care or resources provided, adequately distinguishing high and low resource use. If face validity is the only validity addressed, it must be assessed systematically. Reliability and validity testing can be demonstrated at the measure score or the data element level.

The committee's discussion of the considerations for demonstrating the reliability and validity of a resource use measure and the risk adjustment approach aligned very closely with the guidance presented in the *2011 NQF Testing Task Force Report*. The cumulative experience of the multiple TAPs and the Resource Use Steering Committee demonstrated that resource use measure developers are at various levels of measure testing sophistication. Measures submitted as resource use national consensus standards must demonstrate reliability and validity at the threshold for meeting the scientific acceptability criteria. To

balance the developer burden of testing for the initial evaluation of resource use measures with providing the experts the information needed to make a valid conclusion about reliability and validity, the TAPs and steering committee agreed that the scope of testing may be on a relatively small scale for initial endorsement. The committee agreed further analysis by all developers would be required to support continued endorsement at the time of review in order to maintain NQF endorsement.

RELIABILITY

The NQF evaluation criteria states that reliability testing should demonstrate that the 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, or that the measure score is precise. The committee agreed that developers can demonstrate that the measure score is precise by demonstrating an adequate ratio of signal to noise, or how well one can confidently distinguish the performance of one physician from another.²¹ The signal is ability of the measure to identify real differences in performance, whereas the noise is attributed to measurement error. Demonstrating reliability in this context relies on three major drivers: sample size; differences among physicians; and random variation in the measure scores, or measurement error.²² To demonstrate reliability of a resource use measure relying on administrative claims data, developers may focus on precision of the measure score or validity of the data elements.²³

Reliability at the data element level of resource use measures submitted to this project relied on administrative claims, and by virtue of their design as coded programs were repeatable. However, the committee clarified that while coded programs may be repeatable at the data element level, measure developers need to demonstrate adequate validity testing at the data element level.

Reliability of resource use measures at the measure score level needs to demonstrate that

the measure score is precise. Providing confidence intervals in measure reporting does not sufficiently demonstrate reliability of the measure.

NQF does not prescribe what tests of reliability could be performed, specific thresholds for results, or how to weigh the results of this testing since an evaluation should account for the context of the test, measure, and the data source. The evaluation should incorporate both empirical evidence and expert judgment to evaluate whether the specific measure under evaluation by the Committee has sufficiently demonstrated reliability through the measure submission.

Data element reliability

Discussion of data element reliability was driven by the fact that the submitted resource use measures relied on administrative claims data. Administrative claims provide accessible information on the processes of care and can generally be obtained as a byproduct of the care process. While administrative claims data reduces measure error due to manual chart abstraction and transcription, developers cannot rely on the administrative claims to capture patient clinical characteristics accurately without proper data element validity testing. Claims data provide only limited clinical information, lack detail in determining patient health severity, and are subject to variation in coding processes by the accountable entities. The committee agreed that these concerns span measures of quality and resource use and are not limited to the measures currently under evaluation.

Measure score reliability

Measure developers also performed varying levels of reliability assessments at the measure score level. The committee was interested in assessing the measure's precision or ability to detect signal rather than noise. Measures demonstrated lower levels of measure score reliability assessments including parallel development of episode grouper software and Statistical Analysis System software (SAS) using the same specifications. While these tests demonstrated match rates of more than 99.9

percent, they do not facilitate assessments of the measure score's precision, but rather the precision of the software programming. Further, developers whose measures have been in use attempted to demonstrate the reliability of the observed/expected results (O/E) over time. The Committee suggested other robust methodologies that could be used to demonstrate a high level of reliability, including signal-to-noise ratio analysis using analysis of variance (ANOVA) or intra-class correlation coefficient to demonstrate measure score reliability.

VALIDITY

The NQF criteria state that validity testing must demonstrate that the measure data elements are correct or that the measure score correctly reflects the cost of care or resources provided, adequately distinguishing high and low resource use. Developers must demonstrate measures have undergone sufficient validity testing demonstrating that the measurement approach and results align with the measure intent. If only face validity is addressed, it must be assessed systematically. The Committee recommended that validity testing be demonstrated by correlating measure scores with other valid indicators or by showing that the score produces different results when applied to subgroups known to have differences in resource use. Correct conclusions about resource use can be made when validity tests demonstrate that claims used in the measure accurately reflect information in the charts of a representative sample of patients.

The committee considered that most developers submitting to this project may not have direct access to chart abstracted data. However, additional efforts are strongly recommended to ensure that data elements used to develop the resource use measures are valid. The gold standard approach to determining the validity of data elements based on administrative claims data is to assess the agreement of claims data with source of the data elements (e.g. an authoritative source such as the medical record).²⁴ Since the

entire dataset may not be available for such validation, it is acceptable to apply the resource use measure to a simulated data set that should return known values of the data elements and scores. With either approach, when the results obtained for the resource use measure do not match known values in the simulated data set or the abstracted data, an analysis should be conducted to determine the source of error.²⁵ If the error is related to the measure specifications, including code lists, clinical or construction logic, and computer readable programming language, the measure specification should be corrected before submitting for endorsement.

Data element validity

Data element validity testing should provide an analysis of agreement between critical data elements used to construct a measure and another source of the same information considered to be valid.²⁶ The validity testing submitted at the data element level was often weak because there were no comparisons to other independent claims databases or other authoritative data sources (e.g., the patient's medical record).²⁷ In addition, a comparison of the distribution of important variables to the literature would provide a more robust assessment of the validity of the data elements used. Most measures submitted to the project were tested in large administrative claims databases representative of the target population.

Measure score validity

Validity testing at the measure score level often relied on face validity that the measure score was valid based on clinical review and empirical results. The measure score validity can be demonstrated by correlating measure scores with other valid indicators or by showing that the score produces different results when applied to subgroups known to have differences in resource use.²⁸ Developers often demonstrated face validity by describing the distribution of measure score results, outlier status, and type of service. While the committee members accepted this as a minimum threshold

for demonstrating validity, they suggested that more robust methods, including correlating the measure score with other valid indicators, should be applied in future iterations and testing.

Usability

The focus of the usability criteria is to determine whether the measure results are usable for the intended audience. This includes an evaluation of whether the measure is currently in use and the results are being reported for performance improvement and accountability purposes, and whether the results are considered meaningful and useful. For resource use measures, usability also includes the evaluation of whether it has been demonstrated that the measure construct and its components (e.g., risk-adjustment methodology, clinical logic) can be deconstructed to enable transparency and understanding of the measure score.

Resource use measures presented some specific challenges to applying the concepts identified within the usability criterion. For example, the issue of accountability is a charged one. No consensus existed as to who the intended audience of these measures should be—purchasers, the public at large (consumers), health plans, and health plan members are all likely users of this information. It was noted that for the public at large, extra effort would be required to make the reporting of these measure results as clear as possible; ensuring clarity is the focus of consumer-oriented organizations that share data such as these. There was agreement that these measures should not be reported alone, but in the context of quality measures.

Another challenge the TAPs and committees encountered was differentiating between usability and usefulness and determining whether a measure is inherently usable because it is in use. For measures not currently in use, they questioned how usefulness should be demonstrated since there is a lack of experience of the practical application of these measures.

The committee also questioned the usability of measures that are embedded in a complex episode-grouper system in which each individual measure's logic is interwoven and tied to the logic of another measure, which may not be under evaluation. The committee struggled with how to evaluate the usability of a single measure without evaluating the entire grouper system.

The final overarching issue identified within the usability criteria relates to transparency. Many of the TAP and committee members expressed concern over the complexity of certain methodologies used and questioned whether this complexity masks these measures' ability to be transparent. Difficulty understanding how the risk adjustment, severity level assignments, and episode logic work together in a measure may make it difficult for a physician, for example, to understand completely which of his or her patients was included in the costs attributed to them and how the complexity of the patient population has been accounted for in those costs. Some committee members argued that this lack of transparency and understanding of the construction logic affects the ability of the reported measure score to be used and may limit the physician or health plan from identifying how and where to improve scores. Committee members also questioned whether there should be an expectation that these complex measures would require an investment of time to be interpreted and understood. It was pointed out, however, that by using the resource use service categories identified within the measure, action could be taken using the categories in which high costs were most evident (e.g., imaging, outpatient visits).

Feasibility

The feasibility criterion focuses on the extent to which the measure can be implemented with undue burden and identifies any barriers to implementation. The feasibility subcriteria used to evaluate the resource use measures are identical to those used to evaluate quality measures. Because

all of the resource use measures submitted to this project primarily rely on administrative claims data, the subcriteria evaluating the availability of required data via electronic sources and whether the data are routinely generated required very little discussion. The remaining feasibility subcriteria, however, illuminated some important issues related to implementing resource use measures, which often use very complex, sophisticated methodologies to adjust risk and determine episode logic, for example. The TAPs and the committee discussed this issue of complexity for the implementer (and for the users of the results) during their evaluation of susceptibility to errors and inaccuracies. Some members expressed concern that the complexity of the methodologies lends them to user error, most likely on behalf of the programmer who would develop the code to run the measures. This issue may be mitigated by the purchase of a product that is pre-programmed to implement the measure with imported data or the submission of data to an organization that audits, computes the measure, and reports the information back to the user.

The committee acknowledged that some of the measures under evaluation have been in wide use in the commercial sector for many years. The committee also acknowledged the sensitive nature of some of the measures used in markets where financial investments have been made on behalf of purchasers and other users to integrate the measures into their systems for reporting and understanding costs/resource use. Having been in use in the marketplace by health plans and purchasers for many years, these measures often use some proprietary component or are imbedded in sophisticated proprietary products. For product lines that include large episode-grouping tools encompassing many conditions, a user would be required to purchase some or parts of a product suite to run a single episode for diabetes, for example. Because of this, the committee expressed concern that the financial burden on a small group practice or system to purchase proprietary products could be very

significant, thus creating a barrier to measuring resources using NQF-endorsed standards. The context and process by which measures become endorsed as NQF standards requires that the measures meet each of the four criteria and qualify for use for public accountability and performance improvement purposes. While the current use of the measures is taken into consideration within the usability criteria, the committee agreed this does not imply the measure meets the criteria for endorsement.

Harmonization and Best-in-Class

In phase one of this resource use measurement project, the committee agreed that because this is NQF's first effort focused on evaluating resource use measures, identifying "best-in-class" and requiring harmonization among resource use measures would be premature. In the context of resource use measures, similar measures may share the same measure type (e.g., per episode, per capita), or measure the same costs/resources (e.g., actual prices paid vs. standard prices, resource service categories), or address the same population (e.g., people with diabetes). Competing measures would share all of the characteristics previously listed. Among the eight measures recommended for endorsement, there were no competing measures. Recommended measures that were the same measure type were submitted from the same developer and were already harmonized. With the exception of the two non-condition-specific, total-cost-of-care measures (submitted by the same developer and recommended in Cycle 1), which employ different costing methodologies, all recommended measures addressed different populations. Future resource use measure endorsement efforts should explore the potential ways in which harmonization among similar measures might be achieved. Specifically, identifying which measure constructs (e.g., condition-specific episode trigger and end mechanisms, age ranges), if any, could be harmonized for standard measurement is needed in this measurement area. Also, exploring the

implications of harmonization for the resource use measure development community, in which proprietary measure components are common, would be useful as the portfolio of endorsed resource use measures expands.

NEXT STEPS AND FUTURE WORK

Developing measures of efficiency and value is critical to reducing the healthcare cost growth rate. In a first step toward developing efficiency measures, resource use measures must demonstrate they are important to measure, have scientifically acceptable properties, and are usable and feasible. Resource use measures that meet these criteria may be used in conjunction with quality measures to assess efficiency. The steering committee reflected on the mechanism and future work needed to achieve this goal.

The committee recommended future work to define guidance on how to link quality and resource use measures that can be used to assess efficiency. Considerations should include the measure type (e.g., outcome, process, patient experience), measurement period (e.g., single point in time, spanning the measurement year), and the number of quality measures that should be paired with a resource use measure. The committee also considered that quality measures may be used to monitor for underuse on needed care. Continuing its ongoing work in the public sector to develop a public episode grouper for the Medicare population and exploring ways to measure efficiency using a patient-centered approach will be the focus of future NQF efforts in this area.

Future efforts should explore approaches to ensure that providers are benchmarked on cost performance against providers with similar or better quality performance. Benchmarking cohorts

of providers based on quality performance allows for accurate interpretation of cost. Specifically, this method ensures that the resource use performance is compared to only those providers with equal or higher quality performance.^{29,30} When available, the committee agreed that outcome and patient experience of care measures with sufficient reliability (signal to noise) and validity should be selected to assess efficiency.³¹

Efficiency measurement approaches should be patient-centered, building upon previous efforts such as the **NQF Patient-Centered Episodes of Care (EOC) Efficiency Framework**.

This project enabled first-hand experience in reviewing and understanding some of the various approaches for measuring resources and costs in healthcare, and while many lessons were learned, there is still abundant opportunity to apply the principles and recommendations that emerged from this work in future efforts. Additionally, using the recommendations from the committee on improving the evaluation process, updates to the NQF resource use measure submission forms and evaluation criteria will be explored as NQF seeks to continue to enhance the endorsement process for measure submitters and evaluators.

NOTES

- 1 Catlin A, et al., National Health Spending in 2006: A Year of Change for Prescription Drugs, *Health Affairs*, 2008; 27(1):14-29.
- 2 Banks J, et al., Disease and disadvantage in the United States and in England, *JAMA*, 2006;295(17):2037-2045.
- 3 Hoyert DL, et al., Annual summary of vital statistics: 2004, *Pediatrics*, 2006; 117(1):168-183.
- 4 Weiss JE, Mushinski M, International mortality rates and life expectancy: selected countries, Statistical Bulletin—Metropolitan Life Insurance Company, 1999;80(1):13-21.
- 5 Catlin A, et al.
- 6 CMS Grouper Episode for Medicare. Federal Business Opportunities. Obtained from: https://www.fbo.gov/index?s=opportunity&mode=form&id=452b7af0f5752ba0c444853793e8024d&tab=core&_cview=1. Last Accessed August 2011.
- 7 Institute of Medicine (IOM), *Crossing the Quality Chasm: A New Health Systems for the 21st Century*, Washington, DC: National Academies Press; 2001.
- 8 AQA Principles of “Efficiency” Measures April 2006. Available from: <http://www.aqaalliance.org/performance-wg.htm>. Last accessed August 2011.
- 9 Counting Change: Measuring Healthcare prices, costs, and spending. March 2012. Robert Wood Johnson Foundation.
- 10 Atlas RF, The role of PBMs in implementing the Medicare prescription drug benefit, *Health Affairs*, 2004; Jul-Dec;Suppl Web Exclusives:W4-504-15. Review.
- 11 Bahadori K, Doyle-Waters, et al., Economic burden of asthma: a systematic review, *BMC Pulmonary Medicine*, 2009;9(1):24.
- 12 Thorpe KE and Howard DH, The rise in spending among Medicare beneficiaries: the role of chronic disease prevalence and changes in treatment intensity, *Health Affairs*, 2006;25(5):w378.
- 13 Ibid.
- 14 Hornbrook M, Hurtado A, et al., Health care episodes: definition, measurement and use, *Medical Care Research and Review*, 1985;42(2):163.
- 15 Ibid.
- 16 Ibid.
- 17 Ibid.
- 18 Pham HH, Schrag D, et al., Care patterns in Medicare and their implications for pay for performance, *New Eng J Med*, 2007;356(11): 1130-1139.
- 19 Hogan C, Lunney J, et al., Medicare beneficiaries’ costs of care in the last year of life, *Health Affairs*, 2001;20(4):188.
- 20 Hasings C, Mosteller F, Tukey JW, Winsor CP, Low moments for small samples: a comparative study of order statistics, *Annals of Mathematical Statistics*, 1947;18:413-426.
- 21 Scholle, S. H., J. Roski, et al. (2008). “Benchmarking physician performance: reliability of individual and composite measures.” *The American Journal of Managed Care* 14(12): 833.
- 22 Adams J, McGlynn E, et al., Incorporating statistical uncertainty in the use of physician cost profiles, *BMC Health Services Research*, 2010;10(1): 57.
- 23 **NQF Testing Task Force Report** (2011), 13.
- 24 **NQF Testing Task Force Report** (2011), 37.
- 25 Ibid.
- 26 Ibid.
- 27 Ibid.
- 28 Ibid.
- 29 Tompkins CP, Higgins AR, et al., Measuring outcomes and efficiency in Medicare value-based purchasing, *Health Affairs*, 2009;28(2):w251.
- 30 Chung J, Kaleba E, Wozniak G, A Framework for Measuring Healthcare Efficiency and Value. PCPI Work Group on Efficiency and Cost of Care. Aug 2008. White Paper.
- 31 Hussey PS, De Vries H, et al., A systematic review of health care efficiency measures, *Health Services Research*, 2009;44(3):784-805.

APPENDIX A: Specifications for Endorsed Cost and Resource Use Measures 2012

The following tables present the detailed measure specifications for the eight endorsed resource use consensus standards. All information summarized here has been derived directly from the measure developers without modification or alteration (except where measure developers agreed to such modifications) and is current as of August 15, 2011. All proposed voluntary consensus standards are open source, meaning they are fully accessible and disclosed.

Diabetes

- (1557) Relative Resource Use for People with Diabetes (NCQA)

Cardiovascular

- (1558) Relative Resource Use for People with Cardiovascular Conditions (NCQA)

Pulmonary

- (1560) Relative Resource Use for People with Asthma (NCQA)
- (1561) Relative Resource Use for People with COPD (NCQA)
- (1611) ETG Based Pneumonia cost of care (Ingenix/OptumInsight)

Non-Condition Specific

- (1598) Total Resource Use Population-based PMPM Index (HealthPartners)
- (1604) Total Cost of Care Population-based PMPM Index (HealthPartners)

Bone/Joint

- (1609) ETG Based hip and knee replacement cost of care (Ingenix/OptumInsight)

1557: Relative Resource Use for People with Diabetes	
Steward	NCQA
Description	The risk-adjusted relative resource use by health plan members 18-75 years of age who were identified as having diabetes (type 1 and type 2) during the measurement year.
Resource Use Measure Type	Per capita (population- or patient-based)
Data Source	Administrative claims Electronic Clinical Data: Electronic Health Record, Imaging/Diagnostic Study, Laboratory, Pharmacy Paper Records
Level of Analysis	Clinician: Group/Practice Health Plan Integrated Delivery System Population: National, Regional
Costing Method	RRU measures use NCQA's standardized prices. The organization does not report prices based on its contracts and fee schedules, rather it applies a standard price to each service, multiplies it by the number of units of service and reports the resulting standard cost. Using this approach protects proprietary fee schedules and contracts while supporting equitable measure comparison across organizations and across regions without requiring adjustment for levels of service payment.
Tested Population	Commercial; Medicaid; Medicare
Resource Use Service Categories	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; 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
Attribution Approach	Specifications: Using administrative claims data submitted by all organizations, NCQA estimates the expected RRU amounts for each clinical condition for each organization. RRU index amounts are based on the ratio of observed to expected amounts. Results can be assessed at an overall basis, across all members and major clinical conditions, by service category or for a member cohort within a condition. Relative resource use is calculated at the plan-level and no attribution of resource use is made below this level. Attribution of resource use to a particular NCQA submission is based on the product line and reporting type of the plan that the member was enrolled in as of the end of the measure year.

1558: Relative Resource Use for People with Cardiovascular Conditions	
Steward	NCQA
Description	The risk-adjusted relative resource use by health plan members with specific cardiovascular conditions during the measurement year.
Resource Use Measure Type	Per capita (population- or patient-based)
Data Source	Administrative claims Electronic Clinical Data: Electronic Health Record, Imaging/Diagnostic Study, Laboratory, Pharmacy Paper Records
Level of Analysis	Clinician: Group/Practice Health Plan Integrated Delivery System Population: National, Regional
Costing Method	RRU measures use NCQA's standardized prices. The organization does not report prices based on its contracts and fee schedules, rather it applies a standard price to each service, multiplies it by the number of units of service and reports the resulting standard cost. Using this approach protects proprietary fee schedules and contracts while supporting equitable measure comparison across organizations and across regions without requiring adjustment for levels of service payment.
Tested Population	Commercial; Medicaid; Medicare
Resource Use Service Categories	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; 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
Attribution Approach	Specifications: Using administrative claims data submitted by all organizations, NCQA estimates the expected RRU amounts for each clinical condition for each organization. RRU index amounts are based on the ratio of observed to expected amounts. Results can be assessed at an overall basis, across all members and major clinical conditions, by service category or for a member cohort within a condition. Relative resource use is calculated at the plan-level and no attribution of resource use is made below this level. Attribution of resource use to a particular NCQA submission is based on the product line and reporting type of the plan that the member was enrolled in as of the end of the measure year.

1560: Relative Resource Use for People with Asthma	
Steward	NCQA
Description	The risk-adjusted relative resource use by health plan members with asthma during the measurement year.
Resource Use Measure Type	Per capita (population- or patient-based)
Data Source	Administrative claims Electronic Clinical Data: Electronic Health Record, Imaging/Diagnostic Study, Laboratory, Pharmacy Paper Records
Level of Analysis	Clinician: Group/Practice Health Plan Integrated Delivery System Population: National, Regional
Clinical Framework Description	<p>2 Eligibility Criteria: An encounter with a diagnosis; or by multiple asthma medication events. An organization must use both methods to identify the eligible population, but a member only needs to be identified by one to be included in the measure. To identify the eligible population for measurement:</p> <p>Step 1: Health Plan members are identified as having persistent asthma by meeting at least one of the following criteria during both the measurement year and the year prior to the measurement year. Criteria need not be the same across years.</p> <ul style="list-style-type: none"> • At least one ED visit (Tables ASM-A and ASM-B) with asthma as the principal diagnosis, or • At least one acute inpatient claim/encounter (Tables ASM-B) with asthma as the principal diagnosis (Table ASM-A), or • At least four outpatient asthma visits (Table ASM-B) with asthma as one of the listed diagnoses (Table ASM-A) and at least two asthma medication dispensing events (Table ASM-C), or • At least four asthma medication dispensing events (Table ASM-C) <p>Step 2: Since a member can be identified as having persistent asthma using only leukotriene modifiers as the sole asthma medication dispensed in that year, these members must also have at least one diagnosis of asthma (Table ASM-A), in any setting, in the same year as the leukotriene modifier prescription (e.g. measurement year or year prior to the measurement year).</p> <p>Exclusions:</p> <ol style="list-style-type: none"> 1) Active cancer. Exclude members who had at least one face-to-face encounter, in any setting, with any diagnosis of cancer in conjunction with any treatment code (Table RRU-A), during the measurement year. 2) ESRD. Exclude members who had at least one face-to-face encounter with any code to identify ESRD (Table RRU-B), during the measurement year. 3) Organ transplant. Exclude members who had at least one face-to-face encounter, in any setting, with any code to identify organ transplant (Table RRU-C), during the measurement year. 4) HIV/AIDS. Exclude members who had at least two face-to-face encounters in an outpatient or nonacute inpatient setting, or at least one face-to-face encounter in an acute inpatient or ED setting, with any diagnosis of HIV (Table RRU-D), with different dates of service during the measurement year. Refer to Table RRU-E for codes to identify visit type. 5) Members diagnosed with emphysema, COPD, cystic fibrosis or acute respiratory failure (Table ASM-E) on or prior to December 31 of the measurement year.

1560: Relative Resource Use for People with Asthma

Costing Method	<p>RRU measures use NCQA's standardized prices. The organization does not report prices based on its contracts and fee schedules, rather it applies a standard price to each service, multiplies it by the number of units of service and reports the resulting standard cost. The standard pricing approach is based on the following sources of data:</p> <p>Relative values from the Medicare Fee Schedule (Resource-Based Relative Value Scale, or RBRVS)</p> <p>Pharmacy prices published by First Bank Data</p> <p>Inpatient prices based on a model that uses a broad set of averages, representing different local, regional and national health plans across the country.</p> <p>A plan maps a standard price to each service, multiplies it by the number of units of service and reports the resulting standard cost. It then calculates total standard costs for eligible members across different areas of clinical care and aggregates standard costs across services and members to compute the overall relative resource use.</p> <p>All RRU measures report the standard cost for the following categories.</p> <ul style="list-style-type: none"> • Inpatient Facility • Surgery and Procedure • Inpatient Services • Outpatient Services • Evaluation and Management (E&M) • Inpatient Services • Outpatient Services • Diagnostic Laboratory Services • Diagnostic Imaging Services • Pharmacy, Ambulatory
Tested Population	Commercial; Medicaid
Resource Use Service Categories	<p>Inpatient services: Inpatient facility services; Evaluation and management; Procedures and surgeries; Imaging and diagnostic; Lab services; Admissions/ discharges</p> <p>Ambulatory services: Outpatient facility services; Emergency Department; Pharmacy; Evaluation and management; Procedures and surgeries; Imaging and diagnostic; Lab services</p>
Attribution Approach	<p>Specifications: Relative resource use is calculated at the plan-level and no attribution of resource use is made below this level. Attribution of resource use to a particular NCQA submission is based on the product line and reporting type of the plan that the member was enrolled in as of the end of the measure year.</p>

1560: Relative Resource Use for People with Asthma	
Risk Adjustment	<p>The current risk model utilized by NCQA is based on components of the CMS-HCC risk adjustment methodology and accounts for age, gender, and HCC-RRU risk classifications that predict cost variability. For each condition, members are assigned to a clinical cohort category that provides a more specific classification of the condition. A members age, gender, and HCC category determines their risk score (cohort). NCQA then calculates the average per-member per-month (PMPM) cost for each cohort then weights that cost by the total member months within each cohort. Each plan will have its own weight for each cohort since case-mix varies across plans. These weighted cohort PMPMs are then summed across all cohorts to estimate total resource use that would be expected if the “average” plan had the same case-mix as the plan in question. The ratio of the observed- to-expected PMPM utilization indicates the degree to which a plan deviates from expected performance. This is known as indirect standardization.</p>
Stratification	<p>NCQA collects resource measures at the plan level and summarizes across reporting cohorts along the following dimensions:</p> <ul style="list-style-type: none"> Product line (3 levels): Commercial, Medicaid, and Medicare; Reporting type (2 levels): HMO and PPO; Area level (2 levels): national and region; Resource use or utilization (11 levels): inpatient facility, procedure and surgery (inpatient and outpatient), evaluation and management (inpatient and outpatient), laboratory services, imaging services, ambulatory pharmacy, inpatient discharges, emergency department discharges. <p>Stratification of RRU results to control for individual confounding variables is not performed since age, gender and risk variables (comorbidity and disease interactions) that affect healthcare costs are accounted for in the RRU-HCC risk adjustment process. These include age and gender along with one of the 13 assigned HCC-RRU risk categories (e.g. male 18-44 HCC-RRU 1; male 18-44 HCC-RRU 2; male 18-44 HCC-RRU 3; etc...).</p>

1561: Relative Resource Use for People with COPD	
Steward	NCQA
Description	The risk-adjusted relative resource use by health plan members with COPD during the measurement year.
Resource Use Measure Type	Per capita (population- or patient-based)
Data Source	Administrative claims Electronic Clinical Data: Electronic Health Record, Imaging/Diagnostic Study, Laboratory, Pharmacy Paper Records
Level of Analysis	Clinician: Group/Practice Health Plan, Integrated Delivery System, Population: Community, National, Regional
Clinical Framework Description	Members are identified for the eligible population of the measure with a diagnosis of COPD (Table SPR-A) present anytime during the measurement year and who were continuously enrolled for a two year period (the measurement year and the year prior). Codes to Identify COPD: Chronic bronchitis - ICD-9 Diagnosis: 491 Emphysema - ICD-9 Diagnosis: 492 COPD - ICD-9 Diagnosis: 496 Exclusions: 1) Active cancer. Exclude members who had at least one face-to-face encounter, in any setting, with any diagnosis of cancer in conjunction with any treatment code (Table RRU-A), during the measurement year. 2) ESRD. Exclude members who had at least one face-to-face encounter with any code to identify ESRD (Table RRU-B), during the measurement year. 3) Organ transplant. Exclude members who had at least one face-to-face encounter, in any setting, with any code to identify organ transplant (Table RRU-C), during the measurement year. 4) HIV/AIDS. Exclude members who had at least two face-to-face encounters in an outpatient or nonacute inpatient setting, or at least one face-to-face encounter in an acute inpatient or ED setting, with any diagnosis of HIV (Table RRU-D), with different dates of service during the measurement year. Refer to Table RRU-E for codes to identify visit type. 5) Members diagnosed with emphysema, COPD, cystic fibrosis or acute respiratory failure (Table ASM-E) on or prior to December 31 of the measurement year.

1561: Relative Resource Use for People with COPD	
Costing Method	<p>RRU measures use NCQA's standardized prices. The organization does not report prices based on its contracts and fee schedules, rather it applies a standard price to each service, multiplies it by the number of units of service and reports the resulting standard cost. The standard pricing approach is based on the following sources of data:</p> <p>Relative values from the Medicare Fee Schedule (Resource-Based Relative Value Scale, or RBRVS)</p> <p>Pharmacy prices published by First Bank Data</p> <p>Inpatient prices based on a model that uses a broad set of averages, representing different local, regional and national health plans across the country.</p> <p>A plan maps a standard price to each service, multiplies it by the number of units of service and reports the resulting standard cost. It then calculates total standard costs for eligible members across different areas of clinical care and aggregates standard costs across services and members to compute the overall relative resource use.</p> <p>All RRU measures report the standard cost for the following categories.</p> <ul style="list-style-type: none"> • Inpatient Facility • Surgery and Procedure • Inpatient Services • Outpatient Services • Evaluation and Management (E&M) • Inpatient Services • Outpatient Services • Diagnostic Laboratory Services • Diagnostic Imaging Services • Pharmacy, Ambulatory
Tested Population	Commercial; Medicaid; Medicare
Resource Use Service Categories	<p>Inpatient services: Inpatient facility services, Evaluation and management, Procedures and surgeries, Imaging and diagnostic, Lab services</p> <p>Admissions/discharges</p> <p>Ambulatory services: Outpatient facility services, Emergency Department; Pharmacy, Evaluation and management, Procedures and surgeries, Imaging and diagnostic, Lab services</p>
Attribution Approach	<p>Specifications: Relative resource use is calculated at the plan-level and no attribution of resource use is made below this level. Attribution of resource use to a particular NCQA submission is based on the product line and reporting type of the plan that the member was enrolled in as of the end of the measure year.</p>

1561: Relative Resource Use for People with COPD

Risk Adjustment	<p>The current risk model utilized by NCQA is based on components of the CMS-HCC risk adjustment methodology and accounts for age, gender, and HCC-RRU risk classifications that predict cost variability. For each condition, members are assigned to a clinical cohort category that provides a more specific classification of the condition. A members age, gender, and HCC category determines their risk score (cohort). NCQA then calculates the average per-member per-month (PMPM) cost for each cohort then weights that cost by the total member months within each cohort. Each plan will have its own weight for each cohort since case-mix varies across plans. These weighted cohort PMPMs are then summed across all cohorts to estimate total resource use that would be expected if the “average” plan had the same case-mix as the plan in question. The ratio of the observed- to-expected PMPM utilization indicates the degree to which a plan deviates from expected performance. This is known as indirect standardization.</p>
Stratification	<p>NCQA collects resource measures at the plan level and summarizes across reporting cohorts along the following dimensions:</p> <ul style="list-style-type: none"> Product line (3 levels): Commercial, Medicaid, and Medicare; Reporting type (2 levels): HMO and PPO; Area level (2 levels): national and region; Resource use or utilization (11 levels): inpatient facility, procedure and surgery (inpatient and outpatient), evaluation and management (inpatient and outpatient), laboratory services, imaging services, ambulatory pharmacy, inpatient discharges, emergency department discharges. <p>Stratification of RRU results to control for individual confounding variables is not performed since age, gender and risk variables (comorbidity and disease interactions) that affect healthcare costs are accounted for in the RRU-HCC risk adjustment process. These include age and gender along with one of the 13 assigned HCC-RRU risk categories (e.g. male 18-44 HCC-RRU 1; male 18-44 HCC-RRU 2; male 18-44 HCC-RRU 3; etc...).</p>

1611: ETG Based Pneumonia cost of care measure	
Steward	Ingenix/OptumInsight
Description	The measure focuses on resources used to deliver episodes of care for patients with pneumonia. Pneumonia episodes are defined using the Episode Treatment Groups (ETG) methodology and describe the unique presence of the condition for a patient and the services involved in diagnosing, managing and treating pneumonia. A number of resource use measures are defined for pneumonia episodes, including overall cost of care, cost of care by type of service, and the utilization of specific types of services. Each resource use measure is expressed as a cost or a utilization count per episode and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons.
Resource Use Measure Type	Per episode
Data Source	Administrative claims, Other: Both medical and pharmacy administrative service records (claims or encounters) are used to support the measures. Member enrollment span, pharmacy benefit status and age and gender are also required. Provider characteristics, including specialty and unique provider identifier also have importance to support episode grouping, attribution and definition of peers.
Level of Analysis	Clinician: Group/Practice, Individual, Team Facility Health Plan Integrated Delivery System Population: Community, County or City, National, Regional, State
Clinical Framework Description	The pneumonia measure's episodes are defined using the Episode Treatment Group (ETG) methodology. The pneumonia ETG episode building process that supports pneumonia resource use measures has four important steps: Step 1: Identify Records; Assign Record Type and Anchor Records, Classify Diagnoses and Procedures Step 2: Build Episodes from Anchor Records Step 3: Group Non-Anchor Records to Episodes Step 4: Finalize the Episodes (identify co-morbidities and complicating factors, and assign episode severity)
Costing Method	The financial amounts used should be complete and valid, reflecting the total payments related to the service. The financial amount used in resource measurement should reflect all payments for a service, including those made to the provider by payer, patient and other entities. Allowed payments will reflect both the quantity of different services provided as well as the actual unit price of those same services.
Tested Population	Commercial
Resource Use Service Categories	Inpatient services: Inpatient facility services; Inpatient services: Admissions/discharges; 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

1611: ETG Based Pneumonia cost of care measure

<p>Attribution Approach</p>	<p>Guidelines: Both activity-based and population-based approaches should be supported. As a guideline, four different general options for physician episode attribution can be considered to attribute episodes to individual providers – three activity-based and one population-based approach.</p> <p>Approach 1 – Physician Episode Attribution using Professional Service Costs. This attribution approach identifies the responsible physician for an episode as that provider rendering the greatest amount of professional service costs during the episode.</p> <p>Approach 2 – Physician Episode Attribution using Episode Clusters. This attribution approach identifies the responsible physician for an episode as that provider in the peer group owning the greatest number of “clusters” within the episode.</p> <p>Approach 3 – Physician Episode Attribution using Non-Acute Evaluation and Management (E/M) Visits. This attribution approach identifies the responsible physician for an episode as that physician providing the greatest number of non-acute E/M visits within the episode.</p> <p>Approach 4 – Physician Episode Attribution using a Primary Care, Population-based Approach. This approach requires two important steps: 1) Identification of a PCP for each member. 2) Identify the patient’s assigned PCP during the episode period.</p>
<p>Risk Adjustment</p>	<p>ETG first assesses the observed co-morbidities and condition status factors for an episode and the patient’s age and gender. ETG then assigns a weight to each factor found to influence the relative risk of an episode of pneumonia. These weights and factors are condition-specific and were estimated using pneumonia episode results for a large population. The overall severity score for an episode is the sum of these weights for all factors observed. Using the severity score, a severity level is created, with each pneumonia episode assigned to one of four severity levels. The level of severity assigned to an episode is used to support risk adjustment. The risk adjustment approach includes three important steps:</p> <ol style="list-style-type: none"> 1. Compute the observed experience for the provider being measured, across all episodes to be included in the comparison; 2. Compute the experience for peers or a best practice benchmark. Compute this experience at the level of the risk adjustment, in this case base procedure (hip or knee replacement) and severity level. For a peers benchmark, average cost per episode across all peers for the base procedure and severity level can be computed; 3. Compare the observed experience with the risk adjusted peers or benchmark experience – often called the “expected” result. This expected result is adjusted to reflect both the peers/benchmark levels of performance and also the provider’s own case mix of episodes by condition and level of severity. The ratio of observed to expected results can be termed the relative cost ratio and is a risk adjusted measure.
<p>Stratification</p>	<p>ETG stratifies episodes by the intensity of service, or total cost. For a given episode, a severity score is assigned based on demographic factors (gender and age) and the presence of comorbidities and complications. Once a severity score is determined, a severity level, a number between 1 and 4 is assigned based on a table that relates severity levels to severity scores for each ETG. The severity level can then be used to stratify episodes by severity, measured as resource consumption.</p>

1598: Total Resource Use Population-based PMPM Index	
Steward	HealthPartners
Description	The Resource Use Index (RUI) is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.
Resource Use Measure Type	Per capita (population- or patient-based)
Data Source	Administrative claims, Other: Users administrative claims data base, Risk-adjustment Tool, Johns Hopkins ACG System Version 9.0, Standardized costing code table, Total Care Relative Resource Values (TCRRV) specification provided
Level of Analysis	Clinician: Group/Practice; Population: Community
Costing Method	<p>Description:</p> <p>The Total Care Relative Resource Values (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.</p> <p>General Overview of Application:</p> <p>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. The final step is to calibrate the total TCRRVs to the paid ratio between components of care using the paid adjustment factor.</p> <p>www.healthpartners.com/files/56500.pdf OR www.healthpartners.com/tcoc.</p>
Tested Population	Commercial
Resource Use Service Categories	<p>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.); 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.); Durable Medical Equipment (DME)</p>

1598: Total Resource Use Population-based PMPM Index**Attribution Approach**

Guidelines: To determine which members to include in the Total Resource Use measure, there are several options available depending upon your business purpose and unit of measure. If the unit of measure is an entire health plan or employer group, all members will be included in the Total Resource Use measure.

If the unit of measure is a provider and members are required to select a primary care provider, we recommend using the member selected provider.

When the member is not required to select a primary care provider, we recommend the use of an attribution algorithm to identify the member's primary care provider. The measure was tested using this methodology.

1604: Total Cost of Care Population-based PMPM Index	
Steward	HealthPartners
Description	Total Cost of Care reflects a mix of complicated factors such as patient illness burden, service utilization and negotiated prices.
Resource Use Measure Type	Per capita (population- or patient-based)
Data Source	Administrative claims, Other: Users administrative claims data base, Risk-adjustment Tool, Johns Hopkins ACG System Version 9.0,
Level of Analysis	Guideline: Clinician: Group/Practice, Population: Community
Costing Method	The Total Cost of Care considers 100% of health care services in the Total Cost Index and is calculated on a risk-adjusted paid per member per month basis as well benchmarked to a peer group. The paid amount (i.e., allowed) is inclusive of both plan and member liability.
Tested Population	Commercial
Resource Use Service Categories	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.); 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.); Durable Medical Equipment (DME)
Attribution Approach	Guidelines: To determine which members to include in the Total Resource Use measure, there are several options available depending upon your business purpose and unit of measure. If the unit of measure is an entire health plan or employer group, all members will be included in the Total Resource Use measure. If the unit of measure is a provider and members are required to select a primary care provider, we recommend using the member selected provider. When the member is not required to select a primary care provider, we recommend the use of an attribution algorithm to identify the member's primary care provider. The measure was tested using this methodology.

1609: ETG Based hip/knee replacement cost of care measure	
Steward	Ingenix/OptumInsight
Description	<p>This submission is for Hip/Knee Replacement procedure episodes and will cover both measures at the Hip Replacement and Knee Replacement PEGs. The measure focuses on resources used to deliver episodes of care for patients who have undergone a hip or knee replacement and assigns a level of severity (e.g., resources per episode for Knee Replacement, severity level 1 episodes). Hip Replacement and Knee Replacement episodes are initially defined using the Episode Treatment Groups (ETG) methodology and describe the unique presence of the condition for a patient and the services involved in diagnosing, managing and treating the condition. The Procedure Episode Group (PEG) methodology uses the ETG results and further logic to creating a procedure episode that focuses on the Hip Replacement and Knee Replacement component of the care. Procedure episodes identify a unique procedure event as well as the related services performed before and after the procedure including workup and therapy prior to the procedure as well as post-op activities such as repeated surgery and patient follow-up. Together, the ETG and PEG methodologies identify the services involved in diagnosing, managing and treating patients with Hip/Knee Replacements. A methodology to assign a severity level to each episode is employed to group Hip and Knee Replacement episodes by level of risk.</p> <p>Multiple types of resources can be measured for Hip/Knee Replacement episodes, including overall cost of care, cost of care by type of service, and the utilization of specific types of services. Each resource use measure is expressed as a cost or a utilization count per episode and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons.</p>
Resource Use Measure Type	Per episode
Data Source	Administrative claims
Level of Analysis	<p>Clinician: Group/Practice, Individual, Team, Facility, Health Plan, Integrated Delivery System</p> <p>Population: Community, County or City, National, Regional, State</p>
Clinical Framework Description	<p>This measure identifies patients with Hip/Knee Replacement and creates Hip/Knee Replacement episodes of care using the ETG and PEG methodologies described in the ETG_PEG Construction Logic attached in our response to S.2. Each procedure episode of Hip/Knee Replacement is characterized by a PEG Anchor Category ID that specifies the type of procedure; the PEG Anchor Category ID representing Hip Replacement is 71518 and the PEG Anchor Category ID representing Knee Replacement is 71918.</p> <p>An ETG/PEG episode of Hip/Knee Replacement will contain all clinically relevant information related to the procedure. The Hip/Knee Replacement episode clinical framework is defined by the services, or claim lines, that can begin an episode, the primary and incidental diagnosis relationships involved and how records group to an episode, including relative strength of relationship.</p>

1609: ETG Based hip/knee replacement cost of care measure	
Costing Method	The financial amounts used should be complete and valid, reflecting the total payments related to the service. The financial amount used in resource measurement should reflect all payments for a service, including those made to the provider by payer, patient and other entities. Allowed payments will reflect both the quantity of different services provided as well as the actual unit price of those same services.
Tested Population	Commercial
Resource Use Service Categories	Inpatient services: Inpatient facility services, Admissions/discharges Ambulatory services: Outpatient facility services, Emergency Department, Pharmacy, Evaluation and management, Procedures and surgeries, Imaging and diagnostic, Lab services
Attribution Approach	<p>Guidelines: For physician measurement, the primary surgeon is typically attributed the episode, although applications of attribution could be developed to support an alternate approach. Both activity-based and population-based approaches should be supported. As a guideline, four different general options for physician episode attribution can be considered to attribute episodes to individual providers – three activity-based and one population-based approach.</p> <p>Approach 1 – Physician Episode Attribution using Professional Service Costs. This attribution approach identifies the responsible physician for an episode as that provider rendering the greatest amount of professional service costs during the episode.</p> <p>Approach 2 – Physician Episode Attribution using Episode Clusters. This attribution approach identifies the responsible physician for an episode as that provider in the peer group owning the greatest number of “clusters” within the episode.</p> <p>Approach 3 – Physician Episode Attribution using Non-Acute Evaluation and Management (E/M) Visits. This attribution approach identifies the responsible physician for an episode as that physician providing the greatest number of non-acute E/M visits within the episode.</p> <p>Approach 4 – Physician Episode Attribution using a Primary Care, Population-based Approach. This approach requires two important steps: 1) Identification of a PCP for each member. 2) Identify the patient’s assigned PCP during the episode period.</p>

1609: ETG Based hip/knee replacement cost of care measure

Risk Adjustment	<p>The level of severity assigned to an episode is used to support risk adjustment. The risk adjustment approach includes three important steps:</p> <ol style="list-style-type: none"> 1. Compute the observed experience for the provider being measured, across all episodes to be included in the comparison; 2. Compute the experience for peers or a best practice benchmark. Compute this experience at the level of the risk adjustment, in this case base procedure (hip or knee replacement) and severity level. For a peers benchmark, average cost per episode across all peers for the base procedure and severity level can be computed; 3. Compare the observed experience with the risk adjusted peers or benchmark experience - often called the "expected" result. This expected result is adjusted to reflect both the peers/benchmark levels of performance and also the provider's own case mix of episodes by condition and level of severity. The ratio of observed to expected results can be termed the relative cost ratio and is a risk adjusted measure.
Stratification	<p>The severity level can then be used to stratify episodes by severity, measured as resource consumption.</p>

APPENDIX B: Steering Committee

Tom Rosenthal, MD (Co-Chair)

UCLA School of Medicine, Los Angeles, CA

Bruce Steinwald, MBA (Co-Chair)

Independent Consultant, Washington, DC

Paul G. Barnett, PhD

VA Palo Alto Health Care System, Menlo Park, CA

Jack Bowhan

Wisconsin Collaborative for Healthcare Quality, Middleton, WI

Jeptha P. Curtis, MD

Yale University School of Medicine, New Haven, CT

Kurtis S. Elward, MD, MPH

Family Medicine of Albemarle, Charlottesville, VA

William E. Golden, MD

Arkansas Medicaid, Little Rock, AR

Lisa M. Grabert, MPH

American Hospital Association, Washington, DC

Ethan A. Halm, MD, MPH

University of Texas Southwestern Medical Center, Dallas, TX

Ann L. Hendrich, RN, MSN, PhD(c)

Ascension Health, St. Louis, MO

Thomas H. Lee, MD

Partners HealthCare System, Inc., Boston, MA

Jack Needleman, PhD

University of California, Los Angeles School of Public Health

Mary Kay O'Neill, MD, MBA

CIGNA HealthCare, Seattle, WA

David F. Penson, MD, MPH

Vanderbilt University Medical Center, Nashville, TN

Doris Peter, PhD

Consumer Reports, Yonkers, NY

Steve Phillips, MPA

Johnson & Johnson Health Care Systems Inc., Washington, DC

David Redfearn, PhD

WellPoint, Las Vegas, NV Woodland Hills, CA

Jeffrey B. Rich, MD

Mid-Atlantic Cardiothoracic Surgeons Ltd., Norfolk, VA

William L. Rich, III, MD

Northern Virginia Ophthalmology Associates, Falls Church, VA

Barbara A. Rudolph, PhD, MSSW

The Leapfrog Group, Fitchburg, WI

Joseph Stephansky, PhD

Michigan Health & Hospital Association, Lansing, MI

James N. Weinstein, DO, MS

The Dartmouth Institute for Health Policy and Clinical Practice & The Dartmouth-Hitchcock Clinic, Lebanon, NH

Dolores Yanagihara, MPH

Integrated Healthcare Association, Oakland, CA

NQF Staff

Helen Burstin, MD, MPH

Senior Vice President, of Performance Measures

Heidi Bossley, MBA, MSN

Vice President, of Performance Measures

Taroon Amin, MA, MPH

Senior Director

Ashlie Wilbon, RN, MPH

Senior Project Manager

Sarah Fanta

Project Manager

Lauralei Dorian

Project Manager

Evan M. Williamson, MPH, MS

Project Analyst

APPENDIX C: Technical Advisory Panels

Cardiovascular/Diabetes Technical Advisory Panel

Jeptha Curtis, MD, FACC (Co-Chair)

Yale University School of Medicine, New Haven, CT

James Rosenzweig, MD (Co-Chair)

Boston Medical Center and Boston University School of Medicine, Boston, MA

Mary Ann Clark, MHA

Neocure Group, Washington, DC

Constance Hwang, MD, MPH

Resolution Health, Inc., Columbia, MD

Thomas Marwick, MBBS, PhD

Cleveland Clinic, Cleveland, OH

Michael O'Toole, MD

Midwest Heart Specialists, Ltd., Downers Grove, IL

David Palestrant, MD

Cedars-Sinai Medical Center, Los Angeles, CA

Brenda Parker, PharmD

GlaxoSmithKline, Marietta, GA

Katherine Reeder, PhD, RN

University of Kansas School of Nursing, Kansas City, KS

William Weintraub, MD

Christiana Care Health System, Newark, DE

Pulmonary Technical Advisory Panel

Kurtis S. Elward, MD, MPH (Co-Chair)

Family Medicine of Albemarle; Virginia Commonwealth University, Charlottesville, VA

Janet Maurer, MD, MBA (Co-Chair)

American College of Chest Physicians, Northbrook, IL

Gerene Bauldoff, PhD, RN

The Ohio State University, School of Nursing, Columbus, OH

Kathryn Blake, PharmD

Nemours Children's Clinic, Jacksonville, FL

Dale Bratzler, DO, MPH

University of Oklahoma, Health Sciences Center, Oklahoma City, OK

Zab Mosenifar, MD

Cedars Sinai Medical Center, Los Angeles, CA

Linus Santo Tomas, MD, MS

Pulmonary & Critical Care, Medical College of Wisconsin, Milwaukee, WI

Michael Schatz, MD, MS

Kaiser Permanente, Oakland, CA

Richard Stanford, PharmD, MS

GlaxoSmithKline, Research Triangle Park, NC

Bone/Joint Technical Advisory Panel

James Weinstein, DO, MS(Chair)

The Dartmouth Institute for Health Policy; Dartmouth-Hitch Clinic, Lebanon, NH

Mary Kay O'Neill, MD, MBA

CIGNA HealthCare, Seattle, WA

Elizabeth Paxton, MA

Kaiser Permanente, Oakland, CA

John Ratliff, MD, FACS

Thomas Jefferson University, Philadelphia, PA

Catherine Roberts, MD

Mayo Clinic, Phoenix, AZ

Craig Rubin, MD

University of Texas Southwestern Medical School, Dallas, TX

Patricia Sinnott, PT, PhD, MPH

VA health Economics Resource Center, Menlo Park, CA

Cancer Technical Advisory Panel

David Penson, MD, MPH (Chair)

Vanderbilt University Medical Center, Nashville, TN

Rohit Borker, PhD

GlaxoSmithKline, Philadelphia, PA

Steven Chen, MD, MBA

California Medical Association, Camerillo, CA

Timothy Gilligan, MD

Cleveland Clinic Taussig Cancer Institute, Cleveland, OH

Stephen Grossbart, PHD

Catholic Healthcare Partners, Cincinnati, OH

Dwight Kloth, PharmD

Fox Chase Cancer Center, Philadelphia, PA

Louis Potters, MD, FACP

North Shore-Long Island Jewish Health System, New Hyde Park, NY

Jay Schukman, MD

Anthem Blue Cross and Blue Shield, Richmond, VA

John Skibber, MD

University of Texas-MD Anderson Cancer Center, Houston, TX

Louise Walter, MD

University of California - San Francisco, San Francisco, CA

APPENDIX D: Resource Use Measurement Terms

The following resource use measurement terms have been defined based on their use in the context of this project and are important to understanding the concepts in this report.

Attribution—identifying and assigning of a responsible provider or entity (e.g., health plan) for the care delivered for an episode or population.

Benchmarking—the process of comparing the performance of accountable entities with that of their peers or with external best practice results. In developing comparative estimates, results should be risk adjusted for patient-level attributes to support the valid comparisons of these accountable entities.

Carve-outs—the outsourcing of services, such as behavioral health or pharmacy claims, to specialty health plans or claims processing entities or organizations.

Clinical hierarchy—an arrangement of clinical conditions that are ranked according to severity, as “high,” “below,” or “at the same level.” For example, if a patient has COPD and develops bronchitis, COPD would be assigned a greater weight than bronchitis.

Exclusion criteria—criteria applied before a measure is tested in order to remove any individuals with conditions that may skew the final measure score.

Peer groups—the ways in which resource use measures ensure providers and health plans are compared to similar providers and health plans.

Per capita measure—counts all services provided to a person within a specific population, regardless of condition or encounters with system.

Per episode measure—counts resources based on bundles of services that are part of a distinctive

event provided by one or multiple entities (e.g., health services provided associated with an event or series of events for acute myocardial infarction).

Resource use service categories—categories of resource units or services provided care for a patient or population. Resource units are generally identified through claims data and grouped into categories with similar types of claims (e.g., x-rays grouped into imaging category). Categories are generally measured in terms of dollars, but also can include resources not captured on a claim (e.g., nursing hours).

Risk adjustment—a corrective approach designed to reduce any negative or positive consequences associated with caring for patients of higher or lower health risk or propensity to require health services.

Severity levels—pre-determined levels of acuity used to rank and assign patients based on an assessment of the aggregate of their conditions/diagnosis codes.

Standardized pricing—pre-established uniform price for a service, typically based on historical price, replacement cost, or an analysis of completion in the market; removes variation in resource costs due to differences in negotiated prices or geographic differences based on labor or other input costs.

Stratification—division of a population or resource services into distinct, independent strata, or groups of similar data, enabling analysis of the specific subgroups. This type of adjustment can be used to show where disparities exist or where there is a need to expose differences in results.

NATIONAL QUALITY FORUM

1030 15TH STREET, NW, SUITE 800

WASHINGTON, DC 20005

www.qualityforum.org