Tab I

Measure Evaluation Criteria on Resource Use

NATIONAL QUALITY FORUM

TO: Consensus Standards Approval Committee

FR: Sally Turbyville, Senior Director
Ashlie Wilbon, Project Manager
Sarah Fanta, Research Analyst

RE: Proposed NQF Resource Use Measure Evaluation Criteria

DA: October 27, 2010

The CSAC will review the proposed NQF Resource Use Measure Evaluation Criteria during the November 3-4 in-person meeting. These proposed criteria are grounded in the current NQF Evaluation Criteria but expand on language or sub-criteria in order to accommodate resource use measures. This memo includes a summary of the identified themes of public and member comments on the Proposed Resource Use Measure Evaluation Criteria and Steering Committee discussion and response to those comments.

Accompanying this memo are the following documents:

1. **Proposed Resource Use Evaluation Criteria.** There criteria have been updated with track changes to reflect the changes made following Steering Committee discussion of public and member comments on the criteria. These criteria were presented as an appendix to the *Resource Use Measurement White Paper* that serves as the background document and rationale for the updates to the existing NQF Evaluation Criteria. The complete white paper and supplemental materials are available on the project page.
2. **Comment table for Resource Use Evaluation criteria**. While staff has identified themes within the comments received, all comments did not fit within the themes. This table lists all 11 comments received and the NQF/Steering Committee responses.
3. **Resource Use Modules.** This document details the five modules used to describe the resource use measure construct. These modules have guided the development of the Resource Use Measure Submission Items List and Committee discussion of the criteria.
4. **Draft Resource Use Submission Item List**. This list has been developed to align with the components of a resource use measure and allow for adequate evaluation. It is important to note that while the core evaluation criteria (importance, scientific acceptability, usability and feasibility) have not changed, the information requested on the submission form has changed to align with the analytic functions and nuances of resource use measures.

CSAC ACTION REQUIRED

The CSAC is being asked to review and approve the proposed resource use measure evaluation criteria. Once approved by BOD, These criteria will be implemented in the upcoming CDP project to evaluate and endorse resource use measures.

The proposed NQF Resource Use Measure Evaluation Criteria document was posted for public and Member comment as an appendix to the Resource Use Measurement white

paper. Further, as a parallel effort, NQF staff held a series of conference calls with targeted resource use measure developers where the proposed criteria were discussed along with an updated submission item list form which will ensure that the information necessary to evaluate resource use measures is submitted to NQF.

RESOURCE USE MEASURE EVALUATION CRITERIA

Background

In October 2009, NQF initiated a two-phase project aimed at endorsing resource use measures. Prior to the Call for 'Resource Use' Measures in phase two, NQF convened a Steering Committee representing diverse stakeholders in an effort to understand the full implications of this endeavor for NQF and relevant stakeholders. During phase one of this project, the Committee was tasked with identifying the unique attributes of resource use measures that should be considered during their evaluation.

A primary focus of phase one for the Steering Committee was to contribute to and provide guidance on the development of the Resource Use Measurement White Paper. This paper details the resource use measure specification process and identifies the specific issues that present when developing and evaluating these measures, and ultimately informs the NQF Resource Use Measure Evaluation Criteria (Appendix B) that will be used to evaluate the measures for endorsement in phase two.

COMMENTS AND THEIR DISPOSITION

NQF received 11 comments regarding the Proposed Resource Use Measure Evaluation Criteria from six organizations. The majority of comments were favorable, and many suggested areas of improvement. Staff identified several themes for Committee discussion based on the comments received.

THEMES OF COMMENTS AND COMMITTEE RESPONSES

Comment Theme 1: General approach to resource use measurement criteria and evaluation process

- Criteria as proposed is ambiguous and the application of the criteria will be difficult as it allows for too much Steering Committee discretion
- Use cases should be developed to demonstrate how the criteria will be applied
- Using the quality measure evaluation criteria as foundation is not the best approach

Steering Committee Response:

- There was no specific mention of which criteria or sub-criteria were ambiguous; therefore, it was difficult for the Steering Committee to address this concern and further clarify the criteria.
- Staff and Committee agreed that case examples using different types of resource use measures would be a great educational tool for the TAP and Steering Committee members in preparation for their evaluation of submitted measures. The Steering Committee recommended that this tool not become part of a published document

because it would not be able to count for all variations, but rather solely be used for NQF staff and Steering Committee/TAP education. The Steering Committee agreed that using quality measure evaluation criteria as a foundation for resource use evaluation was a logical approach. The criteria are grounded in the evaluation of measurement properties and the adjustment of the submission tool to accommodate resource use measures aligns with adequate evaluation.

Comment Theme 2: Connection to Quality Measures

- Requiring or *preferring* resource measures that were known to be used with existing endorsed quality measures may limit the number of measures submitted or endorsed.

Steering Committee Response:

- The Steering Committee agreed it may be beneficial to change the language and not refer to any *preference* for a resource use measure that is used alongside a quality measure. However, most of the Steering Committee members held that developers or stewards should still provide a list of related quality measures for informational purposes while emphasizing that it is not a requirement.
- The evaluation criteria itself makes no reference to a 'preference' for those resource use measures associated with quality measures; this language is housed in the white paper only.
- With the goal of efficiency measures in mind, a resource use measure linked to quality measures would be a stronger indicator than stand-alone resource use measures. The Committee also recognized that resource use measures used as stand-alone measures can also be useful for some purposes.
- There were concerns among some Steering Committee members that resource use measures may not be readily linked to quality measures at this time. The "preference" or request for a list of associated quality measures imposes a higher bar than is expected for quality measures currently submitted to NQF. Further, this opens the door for the resource measures to be 'judged' based on the quality measures listed. More importantly, if it is a quality measure that is not favored it may be difficult to disentangle the impressions of the quality measure from the value-add of the resource use measure alone.
- Ultimately, the Steering Committee agreed that measure developers should be asked if there are existing quality measures that can be linked to resource use measures on the submission form. This information will be used for informational purposes only and not be a required component of the submission. Importantly, it was noted that this type of information may be useful to end users and implementers of an endorsed measure.

Comment Theme 3: Importance Criteria--*high* or *unexplained* variation requirement

- NQF should not limit the evaluation of resource use measures to only those that examine "…high or unexplained variation…"

Steering Committee Response:

- The Steering Committee agreed to change the word from "high or unexplained variation" to just those areas demonstrating "high impact or variation."

Comment Theme 4: Testing Requirements

- Clarify and provide examples for the reliability and validity testing of resource use measures.
- How should the reliability criteria be defined--is it strictly *repeatability* or is it 'statistical reliability'?
- Is there a *gold standard* for validity testing?
- What are NQF's/Committee's expectations of testing data?

Steering Committee Response:

- The NQF Testing Task Force report was recently approved by the NQF Board of Directors; the information regarding the specific elements of what is expected of submitted measures will be made publicly available shortly. Currently, there is not a prescriptive approach to testing and the evaluation process that enables the Steering Committee to determine how well the testing data submitted meets the criteria. Therefore, the requirements for validity and reliability for resource use measures may be different from quality measures.
- A gold standard for validity testing may not be applicable to resource use measures. The lack of a gold standard, however, should not preclude the necessity to do validity testing.
- The Steering Committee does not wish to prescribe the levels and type of validity and reliability testing. Developers should use their expertise and discretion to determine the rigor and type of testing that should occur. As done with all submitted measures, the Steering Committee will evaluate the testing rigor and results for each measure submitted for endorsement consideration.

Comment Theme 5: Module Components Subject to *Evaluation*

- Request for clarification on the evaluation of the Data Protocol and the Reporting modules.
- Consider the need for flexibility (which specifications do not have) for these two modules. Should these modules be *specified* or submitted as guidance to users?

Steering Committee Response:

- These modules may be too detailed to be required for submission; it may set the bar too high. The Committee agreed that this information should be included as part of the measure developers submission, but not as specifications. This was clarified in the notes of 2a of the evaluation criteria and the white paper, explaining this area requires a greater extent of flexibility in order to accommodate measure implementation needs.
- Some components of the data protocol (e.g., how to address $0 claims) and the reporting module should be included in the specifications and these should be disentangled from the items that will suffice as guidance. NQF staff reviewed this and updated the criteria notes and submission items list accordingly.

Comment Theme 6: Resource Use Composite Measures

- Request that NQF provide more guidance about resource use composite measures
- Conflicting comments about resource use composite measures:
    - Endorsing composite resource use measures should be a priority.

- Composite resource use measures should be handled separately and delayed. A white paper that focuses solely on composite resource use measures to inform a future call for composite measures is necessary prior to NQF implementing a call for these types of measures.

Steering Committee Response:

- The Steering Committee requested that NQF clearly define composite resource use measure and comprehensive resource use measure. These definitions will be added to the white paper and FAQs that will accompany the call for measures.
- During the call, staff explained the NQF definition of a composite resource use measure would be a measure (and specifications) that estimate a provider's total resource use for several or numerous resource use measures. For example, the total resource use for a provider that combines diabetes, heart failure, and acute low back pain resource use. Comprehensive measures, on the other hand, are measures that account for numerous resources service categories (e.g., pharmacy, evaluation and management and emergency department use) within one resource use measure.
- The Steering Committee agrees that for this first effort at evaluating and endorsing resource use measures, it should be limited to single and comprehensive measures, and that more time is required to consider the adequacy of the criteria to evaluate composite resource use measures.

Comment Theme 7: Reporting Module—Attribution, Peer Group

- Several comments expressed preferences for attribution approaches (i.e., multiple versus single attribution), and the appropriate level of attribution depending on the type of measure (i.e., attribution of per capita measures versus episode-based measures).
- Some comments requested that attribution approaches be standardized; others note the need for flexibility depending on the implementation and measure user needs.

Steering Committee Response:

- At this time, there are no known best attribution approaches for resource use measures, but there is a push for measure developers to provide well thought out and tested attribution recommendations and alternatives for the measures they produce. Currently, the submission form implies the submission of a single method of attribution. The form should be flexible to allow for explanation if there are alternative attribution approaches in addition to the primary suggested method. This is useful for the measure developer, but also for the end-user who may intend to use the measure for a specific purpose. The Steering Committee recommended that attribution be submitted for evaluation but strictly as guidance. This will allows users and consumers to have flexibility to meet their needs while still adhering to the endorsed properties of the measure. NQF staff modified the language in the submission form based on the Steering Committee suggestions.

Comment Theme 8: Allow Time-Limited (i.e., un-tested) Resource Use Measures Endorsement for 'Simple' Measures

- Request to reconsider current approach to not consider resource use measures for *time-limited* endorsement. Measures that have been fully developed but not tested may be eligible for time-limited endorsement allowing developers 12 months to submit testing

data for review if the measures are deemed not complex. Resource use measures have been defined by NQF and the Steering Committee as strictly complex and therefore not eligible for time-limited endorsement.
- Some disagreed that all resource use measures are complex.
- By limiting the *Call for Measures* to those that have been tested, the numbers that are eligible endorsed is immediately reduced.

Steering Committee Response:
- The majority of the Steering Committee believed that resource use measures are never really simple; therefore, time-limited endorsement may push forward measures that are not ready for prime time. Further, given the implications of misuse or misinterpretation of resource use measures, most Steering Committee members stressed the need for resource use measures to being tested prior to submission.
- The majority of the Steering Committee agreed that untested resource use measures should not be allowed during this first RU project.

COMMITTEE CONSENSUS ON MEASURE EVALUATION CRITERIA

The Steering Committee reached consensus on the approval of the proposed NQF Resource Use Measure Evaluation Criteria. Following the call, staff reviewed the criteria and made suggested changes.

MEASURE DEVELOPER FEEDBACK

During the white paper comment period, NQF staff reached out to numerous measure developers for input on the items that will be included in the new online submission tool for resource use measures. Staff had previously received input that the current NQF quality measure submission form would not adequately accommodate the submission or evaluation or resource use measures and accompanying specifications. Overall, the measure developers were supportive of the proposed submission items list and requested minor modifications and clarifications to the form. Measure developers also requested the need for examples within the submission form in order to better guide their responses, as well as making the language clearer. There was general consensus that the evaluation criteria were well aligned with the new submission form. Based on discussions with select measure developers thus far, staff anticipates measure submissions during the Call for Measures from Prometheus, National Committee for Quality Assurance (NCQA), Ingenix and American Board Medical Specialties (ABMS).

NEXT STEPS

In parallel NQF efforts, the evaluation criteria is being reviewed and updated for clarity. Once approved, these changes will be integrated into the resource use evaluation criteria. Upon approval, NQF will post the NQF Resource Use Measure Evaluation Criteria to the website. Targeted outreach and educational efforts, including Measure Developer Webinars will be conducted by NQF staff to ensure that the measure developer community is aware of and understands the criteria and rationale for the changes specific to resource use measures. NQF Staff will also provide education and training to Steering Committee and

Technical Advisory Panel members that will be evaluating resource use measures throughout the project.

NQF Memo: Do not cite, quote, or circulate

Proposed Resource Use Measure Evaluation Criteria (v2)

Background

The resource use measure evaluation criteria is grounded in the standard NQF evaluation criteria, keeping the four major criteria (importance, scientific acceptability, usability, and feasibility) in place, but modifying the subcriteria as appropriate to reflect the specific needs of resource use measure evaluation. The notes for the subcriteria have also been updated to provide specific guidance around meeting the criteria for resource use measures, including appropriate data analysis methods and clarification of concepts.

How to read this document*:*

- ***Bold italicized text*:** Additions and substitutions to the original NQF evaluation criteria to accommodate resource use measures are noted by the ***bold italicized text***. Un-bolded black font represents the original NQF Evaluation Criteria verbatim.
- **Red track-changes text:** Updates to the comment version of the Resource Use Evaluation Criteria resulting from Steering Committee discussion of public and member comments have been made using track changes.
- **Blue text:** Changes made to the criteria based on the recommendations in the *Testing and Evidence Task Force* reports are in blue.

Proposed Resource Use Measure Evaluation Criteria
Conditions for Consideration
A. The measure steward is a governmental organization or a Measure Steward Agreement is signed. B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. C. The intended use of the measure includes both public reporting and quality improvement. D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. ***Based on existing NQF policy, complex measures are not eligible or time-limited endorsement. Resource use measures are complex and therefore must be fully tested at the time of submission.***
If all four conditions for consideration are met, candidate measures are evaluated for their suitability based on four sets of standardized criteria: importance to measure and report, scientific acceptability of measure properties, usability, and feasibility. Not all acceptable measures will be strong-or equally strong- among each set of criteria. The assessment of each criterion is a matter of degree; however, all measures must be judged to have met the first criterion, importance to measure and report, in order to be evaluated against the remaining criteria.

Proposed Resource Use Measure Evaluation Criteria
1. Importance to measure and report
Resource use measures will be evaluated based on the extent to which the specific measure focus is important to making ***significant contributions toward understanding healthcare costs*** for a specific high-impact aspect of healthcare where there is ~~***unexplained***~~ variation or a demonstrated high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, ~~***high or unexplained***~~ ***variation in resource use [current and/or future]***, severity of illness, and patient/ societal consequences of poor quality) or overall poor performance.
1a. The measure focus addresses: • Specific national health Goal/Priority identified by the Partners of the NQF convened National Priorities Partnership: OR • Demonstrated high-impact aspect of healthcare[1] (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality). 1b. Demonstration of ***resource use or cost*** problems and opportunity for improvement, i.e., data[1] demonstrating ~~***unexplained***~~ variation[2] in the ***delivery*** of care across providers and/or population groups (disparities in care). ~~1c. The measure focus is: an outcome (e.g., morbidity, mortality,function, health-related quality of life) that isrelevant to, or associated with, a nationalhealth goal/priority, the condition, population,and/or care being addressed; OR if an intermediate outcome, process,structure, etc., there is evidence thatsupports the specific measure focus as follows: --Efficiency[3] – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality. IOM Quality domains:• Effectiveness• Efficiency• Equity• Patient-centered• Safety• Timeliness~~ 1d. The purpose/objective ***of the resource use measure (including its components)*** and the construct for ***resource use/costs*** are clearly described. 1e. ***The resource*** ~~***units***~~ ***use service categories*** ***(i.e., types of resources/costs)*** that are included in ***the resource use measure*** are consistent with and representative of the conceptual construct represented by the measure. Whether the ***resource use*** measure development begins with a conceptual construct or a set of ***resource*** ~~***units***~~***service categories,*** the ~~***units***~~ ***service categories*** included must be conceptually coherent and consistent with the purpose.

Proposed Resource Use Measure Evaluation Criteria
2. Scientific acceptability of the measure properties
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about ***the cost or resources used to deliver care.***
2a. Reliability 2a1. The measure is well defined and precisely specified[4] so that it can be implemented consistently within and across organizations and allow for comparability. EHR measure specifications are based on the quality data set (QDS).[5] 2a2. Reliability testing[6] demonstrates that the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or the measure score is precise. 2b. Validity 2b1. The measure specifications are consistent with the evidence presented to support the focus of measurement under criterion ~~1c~~***1b***. The measure is specified to capture the most inclusive target population indicated by the evidence and exclusions are supported by the evidence. 2b2. Validity testing[7] demonstrates that the measure data elements are correct and/or the measure score correctly reflects ***the cost of care or resources provided,*** adequately distinguishing ***high*****er*** ***and low*****er*** ***cost or resource use.*** 2b3. Exclusions are supported by the clinical evidence[8] otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted with the exclusion; AND – Measure specifications for scoring include computing exclusions so that the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion); AND – If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent[10] (e.g., numerator category computed separately, denominator exclusion category computed separately). 2b4. For outcome measures and other measures (e.g., resource use) when indicated: --an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care[9,11] and has demonstrated adequate discrimination and calibration OR rationale/data support no risk adjustment/stratification. 2b5. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful[12] differences in

performance, OR there is evidence of overall less than optimal performance.

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

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

Proposed Resource Use Measure Evaluation Criteria
3. Usability
Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and are likely to find them useful for decision-making ~~***Usefulness of resource use measures are in the context of quality.***~~
3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting[13] (e.g., focus group, cognitive testing) and informing quality improvement[14] (e.g., quality improvement). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement. 3b. The measure specifications are harmonized with other measures and are applicable to multiple levels and settings.[15] 3c*. Review of existing endorsed measures and measure sets that demonstrate that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality and ***resource use*** for a particular condition or aspect of healthcare, is a more valid or efficient way to measure). ~~***3c. List NQF-endorsed quality measures known to have been used alongside the resource use measure.***~~ 3d. Data ***and result*** detail are maintained such that the ***resource use measure, including the clinical and construction logic for a defined unit for measurement,*** can be decomposed to facilitate transparency and understanding. 3e. Demonstration (through pilot testing or operational data) that the ***resources use*** measure achieves the stated purpose/objective[16].

**This existing criterion was added back to the proposed resource use criteria after the public and member comment period.*

Proposed Resource Use Measure Evaluation Criteria
4. Feasibility
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement[17].
4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. 4b. The required data elements for the resource use measures are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified, and clinical data elements are specified for transition to the electronic health record. 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity. 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified. 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

Notes for Proposed Resource Use Evaluation Criteria

Notes for Importance

1. Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing, or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality or ***performance*** problem.
2. ***Findings from peer reviewed literature review, empirical data are examples of acceptable information that can be used to justify importance and demonstrating ~~unexplained~~ variation. It is the proof of the measure's concept that enables the Committee to determine if the measure is valid in addressing this concept.***
3. Efficiency of care is a measurement construct of cost of care or resource utilization associated with a specified level of quality of care. ***It is a multi-dimensional concept that includes inputs and outputs, and specifically the amount of resources used (the inputs) and the degree of quality achieved (output)—resource use measures alone do not capture efficiency but are a building block of efficiency: Efficiency = fx (quality, resource use).*** Efficiency might be thought of as a ratio, with quality as the numerator and cost as the denominator. As such, efficiency is directly proportional to quality, and inversely proportional to cost. ~~Efficiency of care is a measurement construct of cost of care or resource utilization associated with a specified level of quality of care.~~ It is a measure of the relationship of the cost of care associated with a specific level of performance measured with respect to the other five IOM aims of quality. ~~Efficiency might be thought of as a ratio, with quality as the numerator and cost as the denominator. As such, efficiency is directly proportional to quality, and inversely proportional to cost.~~ (NQF's Measurement Framework: Evaluating Efficiency Across Episodes Of Care; based on AQA Principles of Efficiency Measures (http://www.aqaalliance.org/files/PrinciplesofEfficiencyMeasurementApril2006.doc).

Notes for Scientific Acceptability

4. Well defined***, complete*** and ***precise specifications for resource use measures include ~~each of the five~~ three of the specification modules~~: (i.e. data protocol,~~ measure clinical logic and method, measure construction logic, and adjustments for comparability~~,~~as relavant to the measure. Data protocol steps are critical to the reliability and validity of the measure; specifications must be detailed enough such that users can execute the necessary steps to implement the measure. Further, additional sub-functions within the data protocol and measure reporting modules may require precise specifity as indicated on the submission form and as appropriate to the submitted measure. ~~and reporting~~***). ***To allow for flexibility of measure implementation, clear guidance from the measure developer is required at time of measure submission on those data protocol and measure reporting steps that are not specified with the measure; this guidance will be reviewed for adequacy by the review Committees. For those modules and analytic functions that are required in the submission form that the measure developer deems as not relevant or available, justification for and implications of not specifying those steps is required***. Specifications should also include the identification of target population to whom the measure applies, identification of those from the target population who achieved the

specific measure focus (i.e. target condition, event) measurement time window, exclusions, risk adjustment, definitions, data elements, data source and instructions, sampling, scoring/computation. ***The resource use measure submission form is the platform through which this information is submitted.***

5. EHR measure specifications include data type from the QDS, code lists, EHR field, measure logic, original source of the data, recorder, and setting.
6. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest, split-half reliability**.** Reliability testing of the measure score addresses precision of measurement (e.g. signal-to-noise) ***Reliability for resource use measures should be demonstrated for each of the modules*** ~~***(data protocol methodology, clinical logic and measure construction, stratification, risk adjustment, and costing methodology)***~~***. For those steps not included in the*** ~~***specifications***~~***reliability testing, justification for and implications of not*** ~~***specifying those functions***~~***addressing those steps is required.***
7. Validity testing applies to both the data elements and the computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to testing hypotheses that the measures scores indicate ~~quality of care~~ ***resource use,*** e.g., measure scores are different for groups known to have differences in ~~quality~~ ***resource use*** assessed by another valid ~~quality~~ ***resource use*** measure or method; correlation of measure scores with another valid indicator of ~~quality~~ ***resource use*** for the specific topic; or relationship to conceptually related measures ~~(e.g., scores on process measures to scores on outcome measures).~~ Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish ***higher*** from ***lower resource use or costs***. ***The scoring/aggregation and weighting rules used during measure scoring and construction are consistent with the conceptual construct. If you use differential weighting it should be justified. Differential weights are determined by empirical analyses or a systematic assessment of expert opinion or values-based priorities. This is in addition to weighting the pricing methodology introduces, if any, should be addressed. Validity testing for resource use measures should demonstrate validity for each module or the entire measure score.*** ~~***[(clinical logic and measure construction, risk adjustment, stratification, costing methodology,and adjustments for comparability and reporting (including attribution, peer groups, threshold and outliers, benchmarking)]].***~~ ***For those steps not included in the*** ~~***specifications***~~***demonstration of validity, justification for and implications of not*** ~~***specifying***~~ ***addressing those steps is required.***
8. Examples of evidence that exclusion distorts measure results include, but are not limited to: frequency ***or cost*** of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers. ***Some measures may specify the exclusion of some patients, events or episodes that are known or determined to be high cost. For example, a patient with active cancer may be excluded from a COPD resource use measure because cancer is considered to be the dominant medical condition with known high costs. Or an episode that exceeds a specified threshold (e.g., 3 standard deviations from the mean) relative to episodes of the***

same type may be excluded with the recommendation for those high cost episodes to be examined separately. Exclusions must be justified and supported with appropriate evidence on the effect of the exclusions. Testing for resource use measure exclusions should address the appropriate specification steps (i.e ~~data protocol,~~ clinical logic, and thresholds and outliers). For those exclusions not addressed, justification for and implications of not addressing them is required. Exclusions do not include the algorithms used to identify the population or area of measurement (e.g., if the measure examines diabetes, exclusion testing does not include those patients without diabetes).

9. Risk factors that influence outcomes ***or resource use/cost*** should not be specified as exclusions; exclusions for resource use or cost that influence results must be justified.
10. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions. ***If there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion). If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). Patient co-pays or co-amounts should not exclude a service from inclusion or justification to exclude these patients or services should be provided. Specifically, claims for services received by the patient should be included for episode construction or patient identification and resource use or cost assessment even when the patient pays a portion of the claims, unless otherwise justified—all approaches should be transparent.***
11. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.
12. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less than optimal performance may not demonstrate much variability across providers.

Notes on Usability

13. Public reporting and quality improvements ***(including strategies around cost or resource use management)*** are not limited to provider-level measures—community and population measures also are relevant for reporting and improvement.
14. Informing improvement may be facilitated using ***relevant quality improvement initiatives or cost containment strategies***.

15. Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., influenza immunization of patients in hospitals or nursing homes), related measures for the same target population (e.g., eye exam and HbA1c for patients with diabetes), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.
16. ***Pilot testing results should address how and who has used the measure practically and in effecting decisions (e.g concurrent validity testing using correlation analysis).***

Notes on Feasibility

17. All data collection must conform to laws regarding protected health information. Patient confidentiality is of particular concern with measures based on patient surveys and when there are small numbers of patients.

NATIONAL QUALITY FORUM

Comments received on the Proposed Resource Use Measure Evaluation Criteria

	Submitter	Submitter Organization	Comment	Response
1	Thomas James	Humana Inc.	Conditions for Consideration--Part C should be amended to include among the intended use of the measures to add "development of health care delivery systems" in addition to public reporting and quality improvement. This would allow health plans to use this information in steerage programs, tiered networks, or development of ACOs (page 61) Measure Focus--Trying to get efficiency measures to meet "one or more of the other five IOM aims of quality, effectiveness, equity, patient-centeredness, safety and timeliness" will be asking more of efficiency measures than is realistic. There is a tendency to make one set of tools do more than is possible and that will delay the process because of the discussions, and lead to a poorer set of tools (page 63) Criteria 2e "For outcome measures and other measures ...and evidence-based, risk-adjustment strategy is specified"--this proposal will limit the types of measures of resource use to those which have a medical model bias and do not necessarily take into account more global economic impacts or patient-preference/social impact models. This is a weakness in the proposal as it is too limiting. (page 66) Feasibility--point 4.c Exclusions is a great point as it limits extraneous information.'	Criterion C (Conditions for Consideration) is derived from NQF's mission to endorse national consensus standards for measuring performance improvement and public reporting. While Criterion C does not include all possible uses of these measures, it does not preclude the measures from being used for other purposes. Staff and Committee agree that efficiency measures most appropriately fit in into the efficiency IOM domain, and it may be very difficult to identify other domains that apply. Resource use measures alone do not easily fit into any of the five domains. This criterion (1c), has been removed from the resource use evaluation criteria. Evidence-based risk adjustment is not limited to the use of medical models. The submission process allows for the developer to describe the risk-adjustment approach used and demonstrate through reliability and validity testing that the strategy used is appropriate and demonstrates adequate discrimination and calibration.

2	Sam Ho	UnitedHealthcare	Under Evaluation Criteria, we do not believe "Scientific Acceptability" best describes the criterion for "the extent to which the measures, as specified, produce consistent (reliable) and accurate (valid) results about the cost or resources used to deliver care." Scientific acceptability may apply to quality measures based on science (evidence-based medicine). We agree with the attributes of resource use as described, however they do not reflect scientific acceptability. We suggest describing these simply as "Acceptable Measure Properties."	The term 'Scientific Acceptability" is not meant to preclude validity and reliability testing without the presence of evidence-based medicine, though quality measures are expected to provide that information. NQF staff will clarify the distinction in what is expected to meet his criteria for resource use measures. The request for changing the criteria title is noted.
3	Barbara Rudolph	The Leapfrog Group	Item D. In this section, the assumption is made that all resource use measures are complex. I would disagree, it is possible that in at least two domains (per admission, per procedure/condition level) measures would not necessarily be complex. These measures generally might only be attributable to one facility or one provider, would likely be time-limited (eg., inpatient stay, visit or diagnostic test), have clear codes and specifications. I would suggest that these two domains not be considered a priori as complex. [1] Bodenheimer, T., and D. West. Low Cost Lessons from Grand Junction, Colorado (2010) NEJM, 363:15, NEJM.ORG, Oct. 7, 2010.	The majority of the Steering Committee agreed that resource measures are never really simple; therefore, time limited endorsement may push forward measures that are not ready for prime time. Further, given the implications of misuse or misinterpretation most Steering Committee members expressed the importance of the measures being fully tested and that the testing needed to be evaluated. The majority of the Committee agreed, that especially during this first effort, untested resource use measures should not be allowed.
4	Barbara Rudolph	The Leapfrog Group	Criteria 1. Importance to Measure: The addition to the criteria of "high or unexplained" variation is inappropriate. If there is variation it means that costs can be reduced, and at this particular point, it seems that this is significant. Second, the addition of the requirement for the variation to be "unexplained" is inappropriate. For example, a recent	The Steering Committee agreed to change the wording from "high or unexplained variation" to those areas demonstrating "high impact or variation".

			article in the NEJM [1], titled "Low-Cost Lessons from Grand Junction, Colorado"; refuted criticisms of the Dartmouth Atlas for failing to adjust for regional price variation (a form of "explained" variation). This article cites three independent observations confirming that Grand Junction provides low-cost healthcare. Second, the University of Wisconsin County Health Rankings shows that the population in Grand Junction's Mesa County is far less healthy than those of other US counties with high Medicare costs. And finally, that it is number one in Colorado for the quality of clinical care being provided. The lesson is that sometimes the explained variation is the variation you need to change! Please remove the term "unexplained variation"; from all references in the criteria.'	
5	Christine Chen	Pacific Business Group on Health	We find the white paper to be lacking in understanding of the origin of NQF's measure evaluation criteria, which were developed with a primary focus on measures of process quality and outcomes. In particular, these criteria were developed to stem the tide of narrow and highly-specific quality measures that were being brought to NQF in droves prior to any real testing. The situation with resource use measures is quite different. In particular: Such measures have been in use for a long time by payers to monitor care; They need to be standardized quickly to allow for pairing with already endorsed quality measures for efficiency analysis; and They generally can be described and understood as generic models that are not necessarily condition-specific but are amenable to analysis by condition/procedure when needed.	The Steering Committee agreed that using the NQF quality measure evaluation criteria as a foundation for resource use evaluation was a logical approach, as it was applicable across different type of measures. The criteria are grounded in the evaluation of measurement properties, including resource use measures; substantial adjustments, informed by measure developers, have been made to the resource use submission tool (currently in development) to accommodate the nuances of resource use measures while allowing for adequate evaluation. Parallel with the public and member comment period, NQF staff held a series of conference calls to obtain feedback from known resource use measure developers on

			NQF should decide on the strategy it chooses to take in endorsing measures of resource use. Does it make sense to take every conceivable combination of medical condition/procedure and resource type through the endorsement process, or can measures be endorsed in a more generic form that would allow for rapid standardization to take place throughout the industry?'	the proposed criteria and updated submission form; NQF noted overwhelming support for both documents. Input was also obtained form developers during and through the Public Comment platform. NQF has been working diligently with developers to understand the implications of the endorsement process; it is anticipated that NQF will endorse individual measures of resource use that may measure many resource use service categories as part of their measurement approach. The call for measures will request for measure submissions across 18 conditions. NQF staff is working internally and with funders to further define the scope of the Call for measures and the measure types (episode-based, per capita, procedure-specific, etc.) that will be accepted in this initial effort. Further, the Steering Committee has recommended that for some implementation approaches flexibility should be allowed, and that the effort should focus on endorsing the properties of the measure core to a standard implementation.
6	Christine Chen	Pacific Business Group on Health	Appendix B: Considerations for Consideration, Item D It is proposed, "Based on existing NQF policy, complex measures are not eligible for time-limited endorsement. Resource use measures are complex in nature and therefore must be fully tested at the time of submission." Yet the white paper cites no evidence for this statement. We do not agree that all resource	See response to comment #3.

			use measures should be considered complex. Instead, we suggest that two types of resource use measures (i.e., per admission, per procedure/condition level) would not necessarily be complex. These measures are likely to have a narrow focus and apply to single providers, limited timeframes, and fairly clean coding and specifications.'	
7	Christine Chen	Pacific Business Group on Health	A second concern is that the effort to amend the current NQF measure evaluation criteria to fit resource use measures seems forced and unnatural. It is difficult to see how the proposed criteria would be applied in practice, and, given this ambiguity, leaves too much discretion with future Steering Committees regarding whether or not the criteria are met. We would strongly urge the White Paper Steering Committee to develop a small number of "use cases"; that would help to illustrate how the criteria are intended to be applied. We would suggest the development of "use case"; scenarios for each of the following: A simple per capita resource use measure, e.g., emergency room visits per 1,000 population Total annual cost associated with a specific chronic disease, e.g. diabetes An episode-based cost measure, e.g., relative cost for an episode of maternity care (mother + infant), using (a) standardized pricing and (b) actual pricing. Total annual cost of care PMPY for a broad population, e.g., that covered by a particular physician organization NCQA's Relative Resource Use measures.	Because there was no specific mention of which criteria or sub-criteria were ambiguous, it is difficult for the Committee to address this concern and further clarify the criteria. The Steering Committee agreed that using the NQF quality measure evaluation criteria as a foundation for resource use evaluation was a logical approach, as it was applicable across different type of measures. The criteria are grounded in the evaluation of measurement properties, including resource us measures; substantial adjustments, informed by measure developers, have been made to the resource use submission tool (currently in development) to accommodate the nuances of resource use measures while allowing for adequate evaluation. Staff and Committee agree that case examples using different types of resource use measures would be a great education tool for the TAP and Steering Committee members in preparation for their evaluation of submitted measures. The Steering Committee recommended that this tool not become part of a published document because it would not be able to account for all iterations, but rather solely be used for

				NQF staff and Steering Committee/TAP education.
8	Christine Chen	Pacific Business Group on Health	Appendix B: Proposed Criteria for Usability The report should acknowledge a potential unintended consequence of public reporting of cost and resource use measures, namely that the prevailing public attitude that "more is better"; may drive business to inefficient providers. Another unintended consequence is when a provider learns that they are on the high end of efficiency and, as a result, demands higher prices from its payers. NQF will need to decide whether reporting to the general public is necessary, as opposed to reporting to the providers, payers, and purchasers involved in the measurement exercise. Also, Criterion 3c should be changed to read: "List NQF-endorsed quality measures that can be used alongside the resource use measure" As stated earlier, while certainly preferred, we do not agree that this criterion should be an absolute requirement for endorsement.'	NQF Staff and Committee recognize the challenges of public reporting cost and resource use measures. However, the endorsement of measures *intended* for public reporting is a tenet of the NQF mission and also applies for resource use measures. The majority of the Steering Committee agreed that the request for this information should be part of the submission process. However, they emphasized that while a list of existing quality measures that can be linked to resource use measures be requested during the submission, it will be used for informational purposes only and not be a required component of the submission. The white paper will also remove any indication that there is a current *preference* for the Steering Committee review. It was noted that this type of information may be useful by users and implementers of an endorsed resource use measure.
9	Rebecca Zimmermann	America's Health Insurance Plans	Criterion 1(c) states that a measure must demonstrate an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality. It is unclear if the broader definition of measure focus will include quality as well as resource use. We encourage NQF to review resource use measures as standalone measures. When reporting resource use, quality information should be included where measures are available.	This criterion is specific to outcome and intermediate outcome measures and does not apply to resource use measures. It has been removed from the proposed resource use criteria. The Committee has agreed to focus on the core modules (measure construction logic, measure clinical logic, and adjustments for

			As stated under the "Modules" section, NQF requires the submission of reporting mechanisms as part of the resource use measures, including proposed attribution, sample size, etc. If the reporting mechanisms are to be included in the endorsed measure specifications, NQF will need to develop criteria to evaluate the reporting mechanisms. The criteria should contain a requirement to test the measure and reporting mechanisms.'	comparability) as required specifications and allow developers to submit information for the reporting and data preparation modules as guidelines rather than specifications. These guidelines will be reviewed by the Committee for clarity and transparency, but they will not be evaluated as specifications. The resource use measure submission form has also been updated to reflect this distinction for these modules. The final resource use submission form will reflect this decision.
10	HHS/CMS Staff	HHS	How do these criteria apply to a system (e.g., the ETG or MEG system?	For large commercial developers whose episode-based measures are often times a part of a larger cost analysis of a provider's resource use, this project does not aim to endorse the system of measures. Rather, the criteria will be applied to individual measures. Developers will be encouraged to submit measures that can be reviewed, evaluated, and implemented as "stand-alone" measures.
11	HHS/CMS Staff	HHS	Criteria 3d: It is very difficult to decompose these measures; they can be almost too sophisticated for anyone to understand. The word "decompose" may be too strong, and may need some satisfactory explanation or intuitive explanation of how to decompose with so many variables in place.	To assist developers and resource use measure submitters with identifying the pieces of information that should be submitted to demonstrate transparency and clarity of the measure construct, the resource use modules have been used to guide the development of the submission items list which lists in detail the required elements of the resource use measure. The submission of these items in this manner will facilitate the decomposition and clarity of the measure construct.

Resource Use Measure Modules

The NQF Resource Use Steering Committee has identified five resource use measurement sub-systems or modules:

1. **Data protocol**
    - The data protocol module includes analytic steps, like cleaning or aggregating the relevant data, necessary to complete the specifications and produce valid results.
        a. Data type and steps needed to run measure
        b. Data cleaning steps
            i. For example, approaches to deal with $0 claims, rejected claims etc.
        c. Initial inclusion and exclusion criteria
            i. For example, truncation or removal of low or high dollar claims
2. **Measure clinical logic**
    - The measure's clinical logic includes the steps that identify the condition or event of interest and any clustering of diagnoses or procedures. For example, the diagnoses and procedures that qualifies for a cardiac heart failure episode, including any disease interaction, comorbid conditions, or hierarchical structure to the clinical logic of the model.
        a. Basic framework for clustering or assigning codes. Includes the identification of the condition(s) or event(s) of interest (e.g., diabetes and CHF) and what type of clinical codes or other clinical markers will be used to identify the condition(s) or event(s). The framework also describes if comorbidities, interactions, hierarchies or severity levels are considered.
        b. Identification of distinct and homogenous unit for measurement.
            i. The type of resource use measure. The identification of the type of resource use measure—per capita per population or per patient, per episode, …, per procedure.
            ii. Grouping or assigning algorithms. The listing of all the clinical markers that will be grouped to identify the measure's condition(s) or event(s) of interest (e.g., a list of diabetes diagnoses and CHF ICD-9 and DRG codes).
        c. Treatment of co-morbidities & disease interactions. Describes how and which comorbid conditions and disease interactions are captured in the measure.

DRAFT

d. Any hierarchy of codes or condition groups. Includes any necessary mapping of clinical markings into their respective category and the trumping logic among the categories.
e. Any severity level assignments

3. Measure construction logic

- The measure's construction logic includes steps used to cluster, group or assign claims beyond those associated with the measure's clinical logic. For example, any temporal or spatial (i.e., setting of care) parameters used to determine if a particular diagnosis or event qualifies for the measure of interest.
  a. Establishes rules by which claims are assigned or grouped using the clinical logic
  b. Measure trigger and end mechanisms (e.g., an AMI event through 60 days after AMI or January 1 through December 31st of the measurement year)
  c. Eliminates redundancy and overlap, as appropriate. These steps explain how claims (or other data) are assigned when there are related or overlapping measures. For example, a developer or steward that maintains two different measures both examining resource use during the same or overlapping time periods for some aspect of cardiac care, the specifications should detail how claims are assigned amongst these two measures. These steps are also relevant for conditions that are 'related' or overlap, such as cardiac conditions and diabetes.
  d. Links complementary services. Complementary services are those that often occur as a consequence of each other or in sequence. For example, a surgery services and its associated anesthesiology service; or an emergency department visit that is discharged to (or results in) an inpatient stay.
  e. Identify units of resource service (e.g., an inpatient stay, emergency department visit or a unit of pharmacy service by dosage and amount), including details on how to define one unit. Simply stating the service of interest is not sufficient—e.g., for an inpatient stay how is a stay defined? An admission? Discharge? Etc.

4. Adjustments for comparability

- External factors can mingle and affect or confound a measure's result. Confounding occurs if an extraneous factor causes or influences the outcome (e.g., higher resource use) *and* is associated with the exposure of interest (e.g., episode of diabetes).

DRAFT

Measure developers often include steps to adjust the measure to increase comparability of results among providers, employers, and health plans.

a. Risk adjustment. Risk adjustment analytic steps are 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. These analytic steps may adjust measures for co-morbid conditions and other factors that may influence resource use, but which are not accounted for in the clinical logic (e.g., using exclusions) or construction logic.
b. Stratification. Including a stratification approach is important where known disparities exist or where there is a need to expose differences in results so that stakeholders can take appropriate action. Stratification includes arranging or separating resource use results by certain confounding patient or other relevant characteristics.
c. Costing method. Depending on the perspective, users of resource use measures may be interested in the count of services, the actual amount paid, or a standardized price approach, which allows users to compare the use and intensity of health services while holding actual paid amounts constant.

5. Measure reporting

- Once the resource use measures have been estimated, users must consider and identify options concerning the reporting of measure results. Measure developers often include decisions about assigning or attributing results to providers or entities, identifying the relevant peer group, estimating the benchmark or comparative values, setting and managing thresholds values, considering statistical matters, and sharing or reporting the results.

a. Attribution rules. The assignment of care provided decisions for a measures results to a provider (e.g., physician, physician groups) or other entity (e.g., health plan) that quantify how their use of resources.
b. Peer group identification and assignment. Identifies a provider or entity peer group and provides details on how to identify those in the peer group.
c. Benchmarking or comparative estimates. Because of the lack of evidence of the appropriate mix of resources, resource use measurement usually includes an

DRAFT

approach to estimate a benchmark or comparative amount of performance among peers

d. Threshold or outlier decisions. Threshold determinations can include discarding or "Windsorizing" (truncating); applying thresholds or removing outlier providers or measures may provide more context for the values. Outliers can be the result of inappropriate treatment, rare or extremely complicated cases, or coding error. Users often do not completely discard outliers, but rather examine them separately.
e. Reporting with descriptive statistics (e.g., distribution, confidence intervals). Depending on the perspective and whether the measure will be used for internal improvement or public reporting, decisions about which statistics must accompany the resource use measure results are critical.
f. List of quality (process or outcomes) measures with which the resource use measure has been reported alongside or is related. The link to quality is critical to determine an input's value. Resource use measures that are used alongside quality or health outcome measures will be given preference over those that are not. Resource use measures that are used this way are one step closer to the goal of understanding efficiency and the value of care provided.

DRAFT

DRAFT V.3.0 Resource Use Measure Submission Form Items:
This document contains a list of proposed items to support the evaluation of resource use measures

Note: Green highlighted text references the evaluation criterion that will be used to evaluate the submission item

RESOURCE USE DEFINITION

- Resource use measures are broadly applicable and comparable measures of input counts—(in terms of units or dollars)-- applied to a population or population sample
- Resource use measures count the frequency of specific resources; these resource units may be monetized, as appropriate.
- The approach to monetizing resource use varies and often depends on the perspective of the measurer and those being measured. Monetizing resource use allows for the aggregation across resources.

NQF CONDITIONS

Please note: These conditions are standard for all measures submitted to NQF for endorsement. All measures must meet criterion, A, B, C and D in order to be considered for endorsement.

A. The measure steward is a governmental organization or a Measure Steward Agreement is signed.

- Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?
- Please check if either of the following apply
  - Proprietary measure
  - Proprietary complex measure with fees

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. (Measure evaluation criterion B) *

C. The intended use of the measure includes both public reporting and quality improvement.

Purpose:

- Public reporting
- Internal quality improvement

Additional purposes:

- Accountability
- Payment incentive
- Accreditation
- Other

D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided.

- Yes, fully developed and tested

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

SPECIFICATIONS

- **Measure Web Page**
  Do you have a page where current detailed measure specifications can be obtained?
  - Yes, Provide the web page URL:
  - No

DESCRIPTIVE INFORMATION:

- Measure Title
- Brief description of measure *(Including type of score, measure focus, target population, event or diagnosis, time, etc.)*
- If this measure is paired with another measure, please identify the paired measure.
- **National Priority Partnership priority area** *(Select the most relevant)*
  - Patient and family engagement
  - Care coordination
  - Population health
  - Palliative and end of life care
  - Safety
  - Overuse
  - Equitable access
  - Infrastructure support
- **IOM Quality Domain** *(Select the most relevant)*
  - Effectiveness
  - Patient-centered
  - Efficiency
  - Safety
  - Equity
  - Timeliness
- **Consumer Care Need** *(Select the most relevant)*
  - Getting better
  - Staying healthy
  - Living with illness

2. MEASURE SPECIFICATIONS

- **2a1/2b1. Type of resource use measure:**
  - Per capita (population- or patient-based)
  - Per episode
  - Per admission (e.g. hospitalization)
  - Per procedure
  - Other. Describe:_____
- **2a1/2b1. Data Dictionary** (Provide a web page URL or attachment.)
  - URL
  - Attachment
- **2a1/2b1.Code Table** *(Please provide a web page URL or attachment.)*
  - URL
  - Attachment
  -

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

Data protocol

- **2a1/2b1. Data preparation for analysis.**

The measure developer must determine if and which of the data preparation steps listed below are measure specifications or should be submitted as guidelines in Section 2b. Specifications allow for no user options and must be strictly adhered to; guidelines are well thought out guidance to users while allowing for needed user flexibility. If the measure developer determines that the measure data preparations for analysis are better suited as guidelines, please select "NA-See Measure Specification Guidelines," otherwise specifications must be provided.

  - Detail (specify) the data preparation steps (e.g., approaches to deal with $0 claims, rejected claims etc.)
    - Describe rationale for data preparation steps (e.g., approaches to deal with $0 claims, rejected claims etc.)
    - NA-See ***Measure Specification Guidelines*** section
  - Detail initial data inclusion criteria (related to claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim)
    - Describe rationale for Initial inclusion criteria
    - NA-See ***Measure Specification Guidelines*** section
  - Detail initial data exclusion criteria (related to claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim)
    - Describe rationale for exclusion inclusion criteria
    - NA-See ***Measure Specification Guidelines*** section
  - Detail steps associated with missing data (e.g., any statistical techniques used)
    - Rationale for missing data steps
    - NA-See ***Measure Specification Guidelines*** section

- **2b6.Data Source: *(Check all that apply)***
  - Electronic administrative data/claims
  - Pharmacy data
  - Lab data
  - Electronic clinical data
  - Electronic Health/Medical
  - Public health data/vital statistics
  - Record Registry data
  - External audit Special or unique data
  - Management data
  - Organizational policies and procedures
  - Paper medical record/flow-sheet
  - Survey: Provider
  - Survey: Patient
  - Documentation of original self-assessment
- **2b6. Data Source or Collection Instrument**
  - Identify the specific data source or data collection instrument (e.g. name of data base, clinical registry, collection instrument, etc.)
- **2b6. Data Source or Collection Instrument Reference**
  - Please provide a web page or UR or attachment.

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

Measure clinical logic

The measure's clinical logic includes the steps that identify the condition or event of interest and any clustering of diagnoses or procedures. For example, the diagnoses and procedures that qualifies for a cardiac heart failure episode, including any disease interaction, comorbid conditions, or hierarchical structure to the clinical logic of the model. (Some of the steps listed separately below may be embedded in the risk adjustment description, if so, please indicate NA and in the rationale space list 'see risk adjustment details.')

- **1e/2a1/2b1/3d.Clinical framework**
  - Detail the framework for clustering and assigning codes, including the grouping methodology and/or the assignment algorithms (For example, using primary ICD-9 diagnoses in the inpatient setting to identify diagnoses of diabetes)
    - Rationale for cluster, grouping and assignment framework
    - Rationale for cluster, grouping and assignment of specified codes (e.g., diagnosis or procedure codes)
- **1e/2a1/2b1/3d.Comorbid and interactions**
  - Detail the treatment of co-morbidities & disease interactions used
    - Rationale for the treatment of co-morbidities & disease interactions
- **1e/2a1/2b1/3d.Clinical hierarchies**
  - Detail the hierarchy of codes or condition groups used
    - Rationale for the hierarchy of codes or condition groups used
- **1e/2a1/2b1/3d.Clinical severity levels**
  - Describe in detail the method used for assigning severity level
    - Rationale for method used for assigning severity level
- **1e/2a1/2b1/3d.Concurrency of clinical events (that may lead to a distinct measure)**
- **1e/2a1/2b1/3d.**Describe in detail the method used for identifying concurrent clinical events and how to manage them (For example, there may be some diagnoses that are generally included in the submitted measure's clinical logic until it has occurred in both the inpatient and outpatient setting, at which time it constitutes or is a part of a clinical framework for another measure. Specifications and rules on how to manage concurrent clinical events is necessary, if applicable)
  - Rationale for method used

Measure construction logic

The measure's construction logic includes steps used to cluster, group or assign claims beyond those associated with the measure's clinical logic. For example, any temporal or spatial (i.e., setting of care) parameters used to determine if a particular diagnosis or event qualifies for the measure of interest.

- **1e/2a1/2b1/3d.Measure Trigger and End mechanisms**
  - Detail the measure's trigger and end mechanisms (e.g., a trigger could be an AMI event or January 1 of the measurement year, and an end mechanisms could be post 60 days of trigger or Dec 31 of the measurement year)
    - Rationale for the measure's trigger and end mechanisms
- **1e/2a1/2b1/3d. Measure redundancy or overlap**
  - Detail how redundancy and overlap of measures can be addressed. (For example, if this is a measure for cardiac care, detail how claims are assigned to more than one measure examining resource use for cardiac conditions or a known 'overlapping' condition like diabetes, during the same or overlapping time periods.)
    - Rationale for methodology used to address redundancy and overlap of measures

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

- **1e/2a1/2b1/3d.Complementary services**
    - Detail how complementary services have been linked to the measure. (For example, describe how an emergency department visit that result in an inpatient admission is captured.)
        - Rational for linking complementary services.

- **1e /2a1/2b1/3d.Resource Units (*select all that apply*)**
    - **Inpatient services**
        - Inpatient facility services
        - Evaluation and management
        - Procedures and surgeries
        - Imaging and diagnostic
        - Lab services
        - Admits/discharges
        - Labor (hours, FTE, etc.)
        - Other inpatient services, list: _____
    - **Ambulatory services**
        - Outpatient facility services
        - Emergency Department
        - Pharmacy
        - Evaluation and management
        - Procedures and surgeries
        - Imaging and diagnostic
        - Lab services
        - Labor (hours, FTE, etc.)
        - Other ambulatory services , list: _____
    - **Durable Medical Equipment (DME)**
    - **Other services not listed, list: _____**
- **1e/2a1/2b1/3d.Resource Units**
    - Describe method or algorithms to identify resource units, including codes, logic and definitions.
    - Provide specifications in either URL (preferred) or as an attachment:
        - URL
        - Attachment
- **2a1/2b1. Care Setting (check all that apply)**--*provides information on which care settings the measure encompasses.*
        - **Ambulatory Care**
            - Ambulatory surgery center
            - Office
            - Clinic
            - Emergency department
            - Hospital outpatient
        - **Inpatient Care**
            - Hospital
            - Office Hospital
            - Long term acute care hospital
            - Nursing home (NH) /skilled nursing facility (SNF)

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

        - Hospital Outpatient Rehabilitation facility
        - Behavioral health/psychiatric unit
    - **Other setting**
        - Assisted living
        - Hospice
        - Home health
        - Dialysis facility
        - All settings
        - Emergency medical services/ambulance
        - Group homes
    - **Unspecified or "not applicable "**
    - **Other, list:**

2a**. Adjustments for Comparability**

- **2b4/3d. Detail the risk adjustment method**
    - Define risk adjustment variables and describe the conceptual, statistical, or other relevant aspects of the model.
    - Rationale for risk adjustment model used
    - Provide a web URL (preferred) or attachment with the risk adjustment specifications.
        - URL
        - Attachment
- **2b4/3d. Detail the stratification method. (Includes all variables, codes, logic or definitions required to stratify the measure)**
    - Rationale for stratification method
- **2b4/3d. Detail the costing method (Includes source of cost information, steps to capture, apply or estimate cost information)**
    - Rationale for costing method

Measure Reporting.

*The measure developer must determine if and which of the following Measure Reporting functions: attribution approach, peer group, outliers and thresholds, sample size, and benchmarking and comparative estimates, are measure **specifications** or should be submitted as **guidelines** in Section 2b. Specifications allow for **no** user options or flexibility and must be strictly adhered to; guidelines are well thought out guidance to users while allowing for needed user flexibility. If the measure developer determines that the requested specification approach is better suited as guidelines, please select "NA-See Measure Specification Guidelines," otherwise specifications must be provided. (Not all the measure reporting approaches have this option.)*

- **2a1.Detail attribution approach**
    - Detail the attribution rules used for attributing costs to providers (E.g., a proportion of total measure cost or frequency of visits during the measure's measurement period)
    - Rationale for the attribution rules
    - NA-See ***Measure Specification Guidelines*** section
- **2a1.Identify and define peer group**
    - Identify the peer group and detail how peer group is identified
        - Rationale for identified peer group and approach to identify those that are part of the defined peer group.
        - NA-See ***Measure Specification Guidelines*** section
- **2a1.Detail measure outliers or thresholds**

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

- Detail any threshold or outlier rules and decisions based on measure costs
  - Rationale for the threshold or outlier rules and decisions
  - NA-See ***Measure Specification Guidelines*** section

- **2a1.Detail sample size requirements**
  - Includes rules associated with type of measure—e.g., at least 30 of one measure or at least 30 across all measures that were attributed to a provider.
  - NA-See ***Measure Specification Guidelines*** section
  -
- **2a1.Define benchmarking or comparative estimates**
  - Detail steps to produce benchmarking and comparative estimates
    - Rationale for benchmarking and comparative estimates
    - NA-See ***Measure Specification Guidelines*** section

- **2a1/2b1.Level of Measurement/Analysis/Attribution (Check the level for which the measure is specified and tested-check all that apply.)**
  - **Clinicians**
    - Individual
    - Group
    - Other clinician, list:
  - **Other**
    - Facility/Agency
    - Health plan
    - Integrated delivery system
    - Multi-site/corporate chain
    - Prescription drug plan
  - **Population**
    - National
    - Regional/Network
    - State
    - County or city
  - **Program**
    - Disease management
    - Quality improvement organization
    - Other program, list:
  - **Can be attributed at all levels**
  - **Other not listed, list:**
- **2b5/3d.Type of Score**
  - Proportion (rate)
  - Dichotomous
  - Ratio
  - Frequency distribution
  - Count
  - Continuous variable
  - Composite
- **2b5/3d. Detail Score Estimation**
  - Detail steps to estimate measure score.

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

- **2b5/3d. Interpretation of Score** *(Classifies interpretation of score according to whether higher or lower resource use amounts is associated with a higher or lower score, a score falling within a defined interval, or a passing score, etc)*
    - Describe the interpretation of the score, including the rationale for the type of score.
- **2b5/3d. Describe discriminating results approach**
    - Detail methods for discriminating differences (reporting with descriptive statistics--e.g., distribution, confidence intervals)
        - Methods to Identify Statistically Significant and Practical or Meaningful Differences in Performance *(Type of analysis and rationale)*

Measure Specification Guidelines

This section requested detailed guidance or guidelines from the measure developer or steward to implementers and users for the submitted measure on critical functions of the measure not necessarily included in the specifications. By providing guidance rather than specifications for the components listed below, implementers have more flexibility such that they can meet their need and measurement perspective. These guidelines should demonstrate well thought out approaches or options for users and testing results as appropriate.

Data protocol

- **2a1/2b1.Data steps required for analysis**
    - Detail (specify) the data preparation steps (e.g., approaches to deal with $0 claims, rejected claims etc.) necessary for this measure, if any
        - Describe rationale for data preparation steps (e.g., approaches to deal with $0 claims, rejected claims etc.)
        - NA-See ***Measure Specification*** section
    - Detail initial data inclusion criteria (related to claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim)
        - Describe rationale for Initial inclusion criteria
        - NA-See ***Measure Specification*** section
    - Detail initial data exclusion criteria (related to claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim)
        - Describe rationale for exclusion inclusion criteria
        - NA-See ***Measure Specification*** section
    - Detail steps associated with missing data (e.g., any statistical techniques used)
        - Rationale for missing data steps
        - NA-See ***Measure Specification*** section

Measure Reporting

- **2a1/2b1.Detail attribution approach**
    - Detail the attribution rules used for attributing costs to providers (E.g., a proportion of total measure cost or frequency of visits during the measure's measurement period)
    - Rationale for the attribution rules
        - NA-See ***Measure Specification*** section
- **2a1/2b1.Identify and define peer group**
    - Identify the peer group and detail how peer group is identified
        - Rationale for identified peer group and approach to identify those that are part of the defined peer group.

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

        - NA-See ***Measure Specification*** section
- **2a1/2b1.Detail measure outliers or thresholds**
    - Detail any threshold or outlier rules and decisions based on measure costs
        - Rationale for the threshold or outlier rules and decisions
        - NA-See ***Measure Specification*** section
- **2a1/2b1.Detail sample size requirements**
    - Includes rules associated with type of measure—e.g., at least 30 of one measure or at least 30 across all measures that were attributed to a provider.
        - NA-See ***Measure Specification*** section
- **2a1/2b1.Define benchmarking or comparative estimates**
    - Detail steps to produce benchmarking and comparative estimates
        - Rationale for benchmarking and comparative estimates
        - NA-See ***Measure Specification*** section

1. IMPORTANCE

- **1a. Demonstrated High Impact Aspect of Healthcare (*select all that apply*)**
    - Affects large numbers
    - Frequently performed procedure
    - Leading cause of morbidity/mortality
    - High resource use
    - Severity of illness
    - Patient/societal consequences of poor quality
    - Other: Describe
- **1a. Summary of Evidence of High Impact**
- **1a. Citations and/or Rationale for Evidence of High Impact**
- **1b. Opportunity for Improvement & Measure Intent**
    - **1d.** Describe intent of the measure and its components/ Rationale for analyzing variation in this way
    - **1d/1e.** Briefly explain the intended use and benefits envisioned by use of this measure
    - **1b.**Summary of data demonstrating variation across providers or entities
    - **1b.** Citations for data on variation
- **1b. Summary of Data on Disparities by Population Group**
    - **1b.** Citations for Data on Disparities

2. SCIENTIFIC ACCEPTABILITY (Measurement Properties)

2a2. Reliability (repeatability) Testing. For resource use measure, includes empirical results for specified:

- For each analysis of data protocol, measure clinical logic and construction, measure adjustment for comparability (i.e. risk adjustment, stratification, costing methodologies), and measure reporting (includes attribution, peer groups, threshold and outlier, benchmarking, discriminating differences) for which results are submitted, the following are required:
        - Data Sample (Description of data sample and size)
        - Analytic Methods (Type of reliability test and rationale, method for testing)
        - Testing Results (Reliability statistics, assessment of adequacy in the context of norms for the test conducted)

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

- Finding statement(s)—i.e., is the measure deemed reliable, limitations…?

2b2. Validity Testing: For resource use measure, includes empirical results for specified:

- For each analysis of measure clinical logic and construction, measure adjustments for comparability (i.e. risk adjustment, stratification, costing methodologies), and measure reporting (includes attribution, peer groups, benchmarking; note, testing to support exclusions is requested separately—see below) for which results are submitted, the following are required:
  - Data Sample (Description of data sample and size)
  - Analytic Methods (Type of validity and rationale, method for testing)
  - Testing Results (validity statistics, assessment of adequacy in the context of norms for the test conducted)
  - Finding statement(s)—i.e., is the measure deemed valid, limitations…?

2b3. Testing for Measure Exclusions

- For each analysis of data protocol, clinical logic, and thresholds and outliers, for which results are submitted, the following are required:
  - Summary of evidence/rationale supporting exclusion(s)
  - Data Sample *(Description of data sample and size)*
  - Analytic Method *(Type of analysis and rationale)*
  - Testing Results *(E.g., frequency, variability, sensitivity analyses)*
  - Finding statement(s)—i.e., is the measure biased due to exclusions, limitations…?

3. USABILITY

Meaningful, Understandable, and Useful Information

3a. Current Use

- In Use
- Testing completed

3a. Use in Public Reporting Initiative (Disclosure of performance results to the public at large) *(If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years)*

3a. Use in QI or Other Programs/Initiatives *(If used in quality improvement (QI) or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years)*.

Testing of Interpretability *(Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)*

- **3a.** Data Sample *(Description of data sample and size)*
- **3a.** Methods *(E.g., focus group, survey, QI project)*
- **3a.** Results *(Qualitative or quantitative results and conclusions)*

3b. NQF # and title of similar or related measure: (leave blank if none)

3.c List NQF-endorsed® quality measures that have been or can be reported alongside the resource use measure

- **NQF # and title of quality measures:** *(Leave blank if none)*

NQF DRAFT: For review purposes only. Do not cite, quote, or circulate.

4. FEASIBILITY

Data Elements Generated as Byproduct of Care Processes

- **4a.** How are the data elements needed to compute measure scores generated?
  - Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)
  - Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)
  - Survey
  - Other, list: _____

Electronic Sources

4b. Are all the data elements available electronically? *(Elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims)*

- Yes
- No

4c. Exclusions

- Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?
  - Yes
  - No

4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences

- Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited.
- If audited, provide results.

4e. Data Collection Strategy

- Describe what you have learned/modified as a result of testing and operational use of the measure regarding data collection, availability of data/missing data, timing and frequency of data collection, patient confidentiality, time and cost of data collection, and other feasibility or implementation issues
- Costs to Implement the Measure *(Costs of data collection, fees associated with proprietary measures)*
- Evidence for Costs
- Business Case Documentation

DRAFT