MAP "WORKING" MEASURE SELECTION CRITERIA INTERPRETIVE GUIDE

NATIONAL QUALITY FORUM

Instructions for applying the measure selection criteria:

The measure selection criteria are designed to assist MAP Coordinating Committee and workgroup members in assessing measure sets used in payment and public reporting programs. The criteria have been developed with feedback from the MAP Coordinating Committee, workgroups, and public comment. The criteria are intended to facilitate a structured thought process that results in generating discussion. A rating scale of *Strongly Agree, Agree, Disagree, Strongly Disagree* is offered for each criterion or sub-criterion. An open text box is included in the response tool to capture reflections on the rationale for ratings.

The eight criteria areas are designed to assist in determining whether a measure set is aligned with its intended use and whether the set best reflects 'quality' health and healthcare. The term "measure set" can refer to a collection of measures--for a program, condition, procedure, topic, or population. For the purposes of MAP moving forward, we will qualify all uses of the term measure set to refer to either a "program measure set," a "core measure set" for a setting, or a "condition measure set." The following eight criteria apply to the evaluation of program measure sets; a subset of the criteria apply to condition measure sets.

FOR CRITERION 1 – NQF ENDORSEMENT:

The optimal option is for all measures in the program measure set to be NQF endorsed or ready for NQF expedited review. The endorsement process evaluates individual measures against four main criteria:

1. **'Importance to measure and report"**–how well the measure addresses a specific national health goal/ priority, addresses an area where a performance gap exists, and demonstrates evidence to support the measure focus;
2. **'Scientific acceptability of the measurement properties'** – evaluates the extent to which each measure produces consistent (reliable) and credible (valid) results about the quality of care.
3. **'Usability'-** the extent to which intended audiences (e.g., consumers, purchasers, providers, and policy makers) can understand the results of the measure and are likely to find the measure results useful for decision making.
4. **'Feasibility'** – the extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measures.

To be recommended by MAP, a measure that is not NQF-endorsed must meet the following requirements, so that it can be submitted for expedited review:

- the extent to which the measure(s) under consideration has been sufficiently tested and/or in widespread use
- whether the scope of the project/measure set is relatively narrow
- time-sensitive legislative/regulatory mandate for the measure(s)
- Measures that are NQF-endorsed are broadly available for quality improvement and public accountability programs. In some instances, there may be evidence that implementation challenges

and/or unintended negative consequences of measurement to individuals or populations may outweigh benefits associated with the use of the performance measure. Additional consideration and discussion by the MAP workgroup or Coordinating Committee may be appropriate prior to selection. To raise concerns on particular measures, please make a note in the included text box under this criterion.

FOR CRITERION 2 – PROGRAM MEASURE SET ADDRESSES THE NATIONAL QUALITY STRATEGY PRIORITIES:

The program's set of measures is expected to adequately address each of the NQS priorities as described in criterion 2.1-2.6. The definition of "adequate" rests on the expert judgment of the Coordinating Committee or workgroup member using the selection criteria. This assessment should consider the current landscape of NQF-endorsed measures available for selection within each of the priority areas.

FOR CRITERION 3 – PROGRAM MEASURE SET ADDRESSES HIGH-IMPACT CONDITIONS:

When evaluating the program measure set, measures that adequately capture information on high-impact conditions should be included based on their relevance to the program's intended population. High-priority Medicare and child health conditions have been determined by NQF's Measure Prioritization Advisory Committee and are included to provide guidance. For programs intended to address high-impact conditions for populations other than Medicare beneficiaries and children (e.g., adult non-Medicare and dual eligible beneficiaries), high-impact conditions can be demonstrated by their high prevalence, high disease burden, and high costs relevant to the program. Examples of other on-going efforts may include research or literature on the adult Medicaid population or other common populations. The definition of "adequate" rests on the expert judgment of the Coordinating Committee or workgroup member using the selection criteria.

FOR CRITERION 4 – PROGRAM MEASURE SET PROMOTES ALIGNMENT WITH SPECIFIC PROGRAM ATTRIBUTES, AS WELL AS ALIGNMENT ACROSS PROGRAMS:

The program measure sets should align with the attributes of the specific program for which they intend to be used. Background material on the program being evaluated and its intended purpose are provided to help with applying the criteria. This should assist with making discernments about the intended care setting(s), level(s) of analysis, and population(s). While the program measure set should address the unique aims of a given program, the overall goal is to harmonize measurement across programs, settings, and between the public and private sectors.

- **Care settings include:** Ambulatory Care, Ambulatory Surgery Center, Clinician Office, Clinic/Urgent Care, Behavioral Health/Psychiatric, Dialysis Facility, Emergency Medical Services - Ambulance, Home Health, Hospice, Hospital- Acute Care Facility, Imaging Facility, Laboratory, Pharmacy, Post-Acute/Long Term Care, Facility, Nursing Home/Skilled Nursing Facility, Rehabilitation.
- **Level of analysis includes:** Clinicians/Individual, Group/Practice, Team, Facility, Health Plan, Integrated Delivery System.
- **Populations include:** Community, County/City, National, Regional, or States. Population includes: Adult/Elderly Care, Children's Health, Disparities Sensitive, Maternal Care, and Special Healthcare Needs.

FOR CRITERION 5 – PROGRAM MEASURE SET INCLUDES AN APPROPRIATE MIX OF MEASURE TYPES:

The program measure set should be evaluated for an appropriate mix of measure types. The definition of "appropriate" rests on the expert judgment of the Coordinating Committee or workgroup member using the selection criteria. The evaluated measure types include:

1. **Outcome measures** – Clinical outcome measures reflect the actual results of care.[1] Patient reported measures assess outcomes and effectiveness of care as experienced by patients and their families. Patient reported measures include measures of patients' understanding of treatment options and care plans, and their feedback on whether care made a difference.[2]
2. **Process measures –** Process denotes what is actually done in giving and receiving care. [3] NQF-endorsement seeks to ensure that process measures have a systematic assessment of the quantity, quality, and consistency of the body of evidence that the measure focus leads to the desired health outcome.[4] Experience of care measures—Defined as patients' perspective on their care.[5]
3. **Cost/resource use/appropriateness measures** –

   a. Cost measures – Total cost of care.

   b. Resource use measures – 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).[6]

   c. Appropriateness measures – Measures that examine the significant clinical, systems, and care coordination aspects involved in the efficient delivery of high-quality services and thereby effectively improve the care of patients and reduce excessive healthcare costs.[7]
4. **Structure measures** – Reflect the conditions in which providers care for patients.[8] This includes the attributes of material resources (such as facilities, equipment, and money), of human resources (such as the number and qualifications of personnel), and of organizational structure

1 National Quality Forum. (2011). The right tools for the job. Retrieved from http://www.qualityforum.org/Measuring_Performance/ABCs/The_Right_Tools_for_the_Job.aspx

2 Consumer-Purchases Disclosure Project. (2011). Ten Criteria for Meaningful and Usable Measures of Performance

3 Donabedian, A. (1988) The quality of care. *JAMA*, 260, 1743-1748.

4 National Quality Forum. (2011). Consensus development process. Retrieved from http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx

5 National Quality Forum. (2011). The right tools for the job. Retrieved from http://www.qualityforum.org/Measuring_Performance/ABCs/The_Right_Tools_for_the_Job.aspx

6 National Quality Forum (2009). National voluntary consensus standards for outpatient imaging efficiency. Retrieved from http://www.qualityforum.org/Publications/2009/08/National_Voluntary_Consensus_Standards_for_Outpatient_Imaging_Efficiency__A_Consensus_Report.aspx

7 National Quality Forum. (2011). The right tools for the job. Retrieved from http://www.qualityforum.org/Measuring_Performance/ABCs/The_Right_Tools_for_the_Job.aspx

8 National Quality Forum. (2011). The right tools for the job. Retrieved from http://www.qualityforum.org/Measuring_Performance/ABCs/The_Right_Tools_for_the_Job.aspx

(such as medical staff organizations, methods of peer review, and methods of reimbursement).[9] In this case, structural measures should be used only when appropriate for the program attributes and the intended population.

FOR CRITERION 6 – PROGRAM MEASURE SET ENABLES MEASUREMENT ACROSS THE PERSON-CENTERED EPISODE OF CARE:

The optimal option is for the program measure set to approach measurement in such a way as to capture a person's natural trajectory through the health and healthcare system over a period of time. Additionally, driving to longitudinal measures that address patients throughout their lifespan, from health, to chronic conditions, and when acutely ill should be emphasized. Evaluating performance in this way can provide insight into how effectively services are coordinated across multiple settings and during critical transition points.

When evaluating subcriteria 6.1-6.3, it is important to note whether the program measure set captures this trajectory (across providers, settings or time). This can be done through the inclusion of individual measures (e.g., 30-day readmission post-hospitalization measure) or multiple measures in concert (e.g., aspirin at arrival for AMI, statins at discharge, AMI 30-day mortality, referral for cardiac rehabilitation).

FOR CRITERION 7 – PROGRAM MEASURE SET INCLUDES CONSIDERATIONS FOR HEALTHCARE DISPARITIES:

Measures sets should be able to detect differences in quality among populations or social groupings. Measures should be stratified by demographic information (e.g., race, ethnicity, language, gender, disability, and socioeconomic status, rural vs. urban), which will provide important information to help identify and address disparities.[10]

Subcriterion 7.1 seeks to include measures that are known to assess healthcare disparities (e.g., use of interpreter services to prevent disparities for non-English speaking patients).

Subcriterion 7.2 seeks to include disparities-sensitive measures; these are measures that serve to detect not only differences in quality across institutions or in relation to certain benchmarks, but also differences in quality among populations or social groupings (e.g., race/ethnicity, language).

FOR CRITERION 8 – PROGRAM MEASURE SET PROMOTES PARSIMONY:

The optimal option is for the program measure set to support an efficient use of resources in regard to data collection and reporting for accountable entitles, while also measuring the patient's health and healthcare comprehensively.

Subcriterion 8.1 can be evaluated by examining whether the program measure set includes the least number of measures required to capture the program's objectives and data submission that requires the least burden on the part of the accountable entitles.

Subcriterion 8.2 can be evaluated by examining whether the program measure set includes measures that are used across multiple programs (e.g., PQRS, MU, CHIPRA, etc.) and applications (e.g., payment, public reporting, and quality improvement).

9 Donabedian, A. (1988) The quality of care. *JAMA*, 260, 1743-1748.

10 Consumer-Purchases Disclosure Project. (2011). Ten Criteria for Meaningful and Usable Measures of Performance.