NATIONAL QUALITY FORUM

May 2011

CSAC GUIDANCE ON QUALITY PERFORMANCE MEASURE CONSTRUCTION

The Consensus Standards Approval Committee (CSAC) has identified some measure construction practices that result in less than optimal quality performance measurement. The issues are directly related to NQF's measure evaluation criteria. In Table 1, the CSAC provides the following guidance on measure construction practices based on the evaluation criteria that should be included in steering committees' evaluations of quality measures being considered for NQF endorsement.

Table 2 is from the Harmonization report appendix (A-1) and provides suggested formats for describing various measure specifications. The intent was to move toward some standardized ways of describing measure specifications so that related and competing measures can be more easily identified and compared.

Table 1. Quality Performance Measure Construction

Considerations for Quality Performance Measure Construction	**Related Criteria/Rationale**
1. Avoid specifying measures so that they can be met primarily through documentation without an evaluation of the quality of the activity (e.g., often satisfied with a checkbox, date, or code). For example: • an assessment completed; • a careplan created; or • an instruction or advice given (teaching, counseling) Alternatively, consider: • specifying the quality/appropriateness of the activity; • including the activity within a more proximal related measure (e.g., incorporate an assessment activity in a measure of a related intermediate outcome or process); • measuring a related desired outcome, or intermediate clinical outcome or process more proximal to the desired outcome	1c. Evidence for the measure focus Measuring desired outcomes and processes and structures with direct evidence of impact on desired outcomes will facilitate the greatest improvements in quality and health. There may be several issues related to the evidence • The evidence may identify specific characteristics or elements of the particular measure focus, such as the elements of effective smoking cessation counseling, or what constitutes an adequate careplan; whereas the submitted measure just requires an indication that counseling was given or a careplan created. • The evidence may be about a specific intervention, treatment, or intermediate clinical outcome that leads directly to desired outcomes (e.g., giving influenza vaccination, Hba1c level), but the measure focus is on something more distal such as performing an assessment or assessment frequency. *Note: Evidence-based prevention screening (e.g., those recommended by the USPSTF)*

Considerations for Quality Performance Measure Construction	Related Criteria/Rationale
	would have the requisite evidence for the linkages to desired outcomes. 2b. Validity Because of either an indirect nature of the evidence or measure specifications that are not directly reflective of the evidence, such measures often will not be validated by a demonstrated association with other measures of quality. For example, high rates on a specific measure of smoking cessation counseling may not be associated with increasing rates on a measure of smoking cessation.
2. Consider the impact of missing data. Generally, missing data should not be specified as an exclusion or implicitly limit inclusion (e.g., percent of patients with a lab value within norms is often specified so that the denominator includes only those patients who had the lab test).	2b. Validity Missing data may be indicative of a quality problem in itself, so excluding those cases may present an inaccurate representation of quality. Systematic missing data (e.g., if poor performance is selectively not reported) also affects validity of conclusions that can be made about quality. 4c. Susceptibility to inaccuracies, errors of unintended consequences Missing data could result in better performance scores.
3. The effect of exclusions for patient preference should be transparent (e.g., separate numerator category, computed and reported denominator exclusions)	2b. Validity As with #1 above, merely indicating that a patient declined a service or intervention does not indicate the quality of the exchange that occurred between the healthcare provider and patient. Exclusions for patient preference (refusal) could be related to quality problems.
4. Measures should be specified with the broadest applicability (target populations, settings, levels of analysis) as supported by the evidence. Consider stratification to compute and report performance results by different subsets of patients.	1c. Evidence for the measure focus Measures should include all patients indicated by the evidence for all applicable settings and levels of analysis. 3. Usability Such measures have broader usefulness and may also promote shared accountability. 5a. Measure Harmonization (and parsimony) The creation of multiple related measures

Considerations for Quality Performance Measure Construction	Related Criteria/Rationale
	increases the likelihood of different specifications with resulting confusion and loss of ability for comparisons across settings, patient populations, etc.
5. It is preferable to measure teaching/counseling from the patient perspective – i.e., intermediate outcomes of knowledge gained (not just a rating scale); or patient experience of care regarding receipt of understandable teaching/counseling. Alternatively, the specific elements of effective teaching/counseling should be included in measures focused on the delivery of teaching/counseling.	2b. Validity The most important aspect of teaching/counseling is whether the patient understands how to manage care, treatment options, and consequences.
6. Exclusions should be supported by the clinical evidence or supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion.	2b. Validity 2b1. Specifications should be consistent with the evidence provided in support of the measure including the specified exclusions. 2b3. Exclusions If incidence of exclusions is rare or evenly distributed across the entities whose performance is being measured, it is unlikely to affect comparative performance scores. 4d. Data collection strategy – numerous exclusions could increase the burden of data collection.
7. Statistical risk models generally should not include factors related to disparities in care	2b. Validity 2b4. Evidence-based risk adjustment or stratification Including factors associated with disparities in statistical risk models obscures quality problems related to disparities. 2c. Disparities If disparities in care have been identified, measure specifications should allow for identification through stratification.
8. Measures should be fully specified including all applicable definitions and codes	2a. Reliability 2a.1 Precise specifications provide the foundation for reliability. Precise specifications also are essential for both

Considerations for Quality Performance Measure Construction	Related Criteria/Rationale
	testing and implementation. Without detailed specifications that include definitions and codes needed for implementation, it's essentially a measure idea rather than a measure.
9. Adults should be identified as age 18 and older (no upper limit); pediatric as under age 18 (no lower limit) ***unless the evidence indicates otherwise***	5a. Harmonization This was agreed on in prior discussions with developers.
10. Do not include reporting classification methods (e.g., stars) in the basic measure construction and specifications	Currently, NQF endorsement does not include presentations/methods for reporting. Different entities than the measure developer may implement and report on the measure. 2b. Validity The measure performance score and reporting methods could have different issues regarding validity.
11. Avoid measures where improvement decreases the denominator population (e.g., denominator – patients who received a diagnostic test; numerator – patients who inappropriately received the diagnostic test. With improvement, fewer will receive the diagnostic test, decreasing the denominator)	3a./3b. Usefulness for public reporting and quality improvement If the denominator changes with improvement, it may be difficult to interpret and compare scores.

Table 2. Measure Specifications, Suggested Format, and Level of Harmonization (Table A-1 in Measure Harmonization report)

Measure Specification (submission item)	**Construction of Measure Specifications**	**Suggested Format and Example** *(intended only to illustrate the suggested format, not a fully specified measure)*
Measure Title (De.1)	Briefly convey as much information as possible about the measure focus and target population—abbreviated description	[target population] who received/had [measure] *Patients with diabetes who received an eye exam*
Brief Description of Measure (De.2)	Briefly describe the type of score (e.g., percentage, proportion, number) and the target population and focus of measurement	[type of score] of [target population] who received/had [measure focus] *Percentage of adult patients with diabetes who received a foot exam (including visual inspection, sensory exam with monofilament, or pulse exam)*
Measure Focus/Numerator Statement (2a.1) Time Window (2a.2)	Describe the measure focus—cases from the target population with the target process, condition, event, or outcome based on the evidence. If the time frame is different than for identifying the target population, it should be specified.	Patients in the target population who received/had [measure focus] {during [time frame] if different than for target population} *Patients in the target population who received a foot exam including visual inspection, sensory exam with monofilament, or pulse exam*
Measure Focus/ Numerator Details (2a.3)	**Codes:** For measures based on a coded data set, identify the code set, the specific codes, and descriptors for the codes. **Details:** Definitions and instructions as needed. As a starting point, use specifications that exist in the NQF-endorsed measures database or Quality Data Set (QDS) when available.	**Codes:** [concept] [code set] [number or range of numbers] *Reintubation procedure* *ICD-9-CM:* *96.04 Insertion of endotracheal tube OR* *96.70 Invasive mechanical ventilation: Unspecified duration OR* *96.71 Less than 96 hours OR* *96.72 For 96 hr or more* **Details:** [concept] definition or instruction *Reintubation procedure* *ICD-9-CM* *96.04 IF one or more days after the major operating room procedure code* *96.70 or 96.71 IF two or more days after the major operating room procedure code* *96.72 IF zero or more days after the major operating room procedure code*
Target Population/ Denominator Statement (2a.5)	Designate the broadest population based on the evidence for which the target process, condition, event, outcome is applicable. The target population should indicate age, setting, and time frame for	Patients [age] with [condition] in [setting] during [time frame] *Patients (age 18-75) with diabetes in ambulatory care during a measurement*

Measure Specification (submission item)	**Construction of Measure Specifications**	**Suggested Format and Example** *(intended only to illustrate the suggested format, not a fully specified measure)*
Time Window (2a.6)	identifying the target population.	*year*
Target Population/ Denominator Details (2a.7)	**Codes:** For measures based on a coded data set, identify the code set, the specific codes, and descriptors for the codes. **Details:** Definitions and instructions as needed. As a starting point, use specifications that exist in the NQF-endorsed measures database or QDS when available.	**Codes:** [concept] [code set] [number or range of numbers] *Heart failure* *ICD-9-CM codes:* *402.01 Malignant hypertensive heart disease with congestive heart failure (CHF)* **Details:** [concept] definition or instruction *For chart abstraction, identify patients with a diagnosis of heart failure on the problem list*
Exclusions from Target Population/ Denominator (2a.8)	Identify patients who are in the target population, but who should not receive the process or are not eligible for the outcome for some other reason, particularly where their inclusion may bias results. Exclusions should be evidence-based.	Patients in the [target population] who [have some additional characteristic, condition, procedure] *Patients with diabetes who have gestational or steroid-induced diabetes*
Exclusion Details (2a.9)	**Codes:** For measures based on a coded data set, identify the code set, the specific codes, and descriptors for the codes. **Details:** Definitions and instructions as needed. As a starting point, use specifications that exist in the NQF-endorsed measures database or QDS when available.	**Codes:** [concept] [code set] [number or range of numbers] *Gestational diabetes ICD9-CM 648.8* **Details:** [concept] definition or instruction
Calculation Algorithm (2a.20)	Describe the calculation of the measure as a flowchart or series of steps.	*1. Identify all discharges for the calendar year (Jan 1-Dec 31)* *2. Identify patients 18 and older at time of discharge (discharge date-birth date)* *3. Identify patients with CHF (ICD-9 codes listed in denominator details)* *4. Exclude patients if . . .*
Technical Sampling (Survey) Methodology (2a.24)	If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey, and guidance on minimum sample size (response rate).	*For chart abstraction, select a random sample of 30 discharges per month*
Data type (2a.25) Level of analysis (2a.33)	Identify those for which the measure is completely specified and tested.	Check the appropriate boxes.

Measure Specification (submission item)	**Construction of Measure Specifications**	**Suggested Format and Example** *(intended only to illustrate the suggested format, not a fully specified measure)*
Data Source or Collection Instrument (2a.26)	Identify the specific data source or data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.	*Outcome and Assessment Information Set (OASIS)* *MedPAR database*
Data Dictionary or Code Table (2a.30)	Provide URL or attachment (if exceeds 2 pages); however, key definitions should be in the submission form numerator and denominator details.	
Stratification Details/ Variables (2a.10)	Provide instructions for calculating the measure by category (e.g., age) including the stratification variables, all codes, logic, and definitions	*Compute overall hospital score and also by race. Identify patients as white, black, Hispanic, and other and compute results for each group.*
Risk-Adjustment Method/ Variables (2a.13)	Identify the method and variables/risk factors (not the details).	[method] [variables/risk factors] *Logistic regression model* *Risk Factors:* *Age* *Functional status* *Prior hospitalization* *Co-morbid conditions of diabetes, CHF, CAD*
Detailed Risk Model (2a.14)	Provide risk model coefficients or equation to estimate each patient's probability for the outcome including coefficients for the variables/risk factors. Provide the codes or definitions for each variable/risk factor. Provide programming language (e.g., SAS code).	Intercept -9.50 Age/10 0.59 BMI/5 -0.07 Cerebrovascular disease 0.43 Chronic lung disease 0.38