TO: eMeasure Feasibility Expert Panel

FR: Reva Winkler and Helen Burstin

DA: October 24, 2012

RE: October 30 eMeasure Feasibility meeting

This memo presents information on NQF's eMeasure Feasibility Assessment project including background, project goals and questions for discussion at the in-person meeting.

**eMEASURE FEASIBILITY ASSESSMENT PROJECT OVERVIEW**

As quality measurement shifts to measures derived from electronic health records (EHRs) to take advantage of timely clinical data in addition to claims data, there is a need for clarity about the assessment of feasibility needed to assure that eMeasures can be used for a range of accountability applications. eMeasures require both precision and results that are reliable and valid. While the concepts of reliability and validity apply equally to measures derived from EHRs and traditional measures, the electronic health record presents additional challenges related to measure assessment:

- widespread EHR data are not yet available for measure development and assessment;
- a lack of standardization of concepts (codes) and terminology across vendor products;
- data elements needed for advanced measures currently may not be feasible to capture in EHRs; and
- the burden of data capture for front line clinicians.

Realizing the promise of EHRs as a tool for quality reporting will rest on the ability of providers, payers, vendors and other users of the information to know that e-Measures provide valid and reliable data. During the public and member comment for NQF's eMeasure Review and Assessment draft proposal in early 2012, several organizations expressed support that eMeasure assessment should incorporate **the feasibility of data capture for the data elements utilized** in addition to reliability and validity assessment. This requirement is significantly hampered by the lack of clarity and definition in the field as to what constitutes feasibility assessment for EHRs and what a reasonable level of feasibility should be for which the benefits exceed the costs?

The goal of this project is to propose a set of feasibility assessment recommendations, as well as a starter set of criteria for eMeasure feasibility that will assess the feasibility of data capture for the data elements utilized in a structured format and address the following considerations:

- Timing of feasibility assessment in the course of measure development (e.g., iterative assessments with development, implementation of fully specified measures)
- Potential differences in feasibility assessment requirements for de novo measures and retooled measures;
- Number and diversity of EHRs systems used for assessment and relation to comparability across EHRs in terms of data feasibility, validity, and reliability; and
- Interrelationship of feasibility and validity assessment of data elements of new measures.

The deliverable from this effort, the eMeasure feasibility assessment guidance report, should provide important guidance that can shape future measure development contracts as well as certification requirements.

**ENVIRONMENTAL SCAN**

To assist the Panel in understanding the current approaches used to assess eMeasure feasibility, NQF conducted an environmental scan of approaches to feasibility assessment from measure developers, providers and EHR vendors.

Summary of the environmental scan will be provided separately.

**DISCUSSION**

Purpose and goals of feasibility assessment

Assessment should be performed to answer important questions and unknowns about the feasibility of individual eMeasures. A standardized approach to feasibility assessment will provide a common set of expectations for assessment of measures during development as well as expectations for stakeholders in the field when an eMeasure has been tested and determined to be feasible.

Discussion questions:

- What are the expectations for assessment of feasibility of data capture for a specified measure?
- Should an eMeasure be tested in its entirety, or should data elements be tested individually to build a catalogue of available, feasible data elements?
- When assessment can demonstrate that an eMeasure is feasible, what should stakeholders know about the measure assessment?
- Is there a difference between the assessment of feasibility of the vendor product and assessment of feasibility of the measure in the real world that may or may not include local IT support?
- What are the resource and capability limits of measure developers, EHR vendors and providers to perform feasibility assessment?

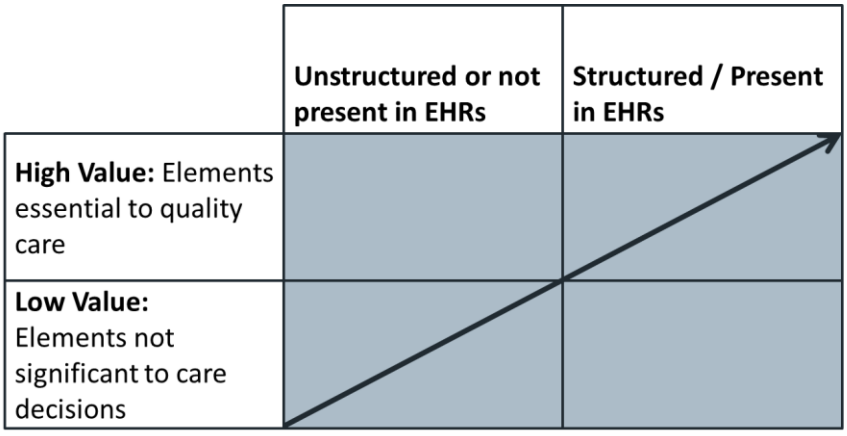
Feasibility in evolution

Both quality measure development and EHR capabilities are rapidly evolving. Stakeholders are demanding more meaningful and progressive measures that address quality concerns important to patients and families. Consumers, purchasers, policy makers and others are looking for measures that assess concepts such as care coordination, patient engagement, as well as longitudinal measures and measures that incorporate patient reported outcomes. The promise of EHRs suggests that these

challenging measures may become more feasible, including measure constructs that are possible only in EHRs. Guidance for feasibility assessment must not hinder the evolution of measures or EHR systems yet must help stakeholders understand what is feasible for their current circumstances.

Discussion questions:

- How should expectations for feasibility assessment accommodate advanced measures to encourage EHRs to evolve in ways that meet the goals of the National Quality Strategy?
- The QDM may not support new types of data or measures – does the QDM lead or follow?
- Should efforts continue to adapt traditional measures to EHRs despite significant feasibility challenges? At what point should efforts to retool traditional measures be abandoned?
- How should feasibility and the value of a measure's data elements be considered? (see ONC diagram below)
- Are there potential levels of feasibility for eMeasures?
    - Level 1- data all in your own machine/network OR measures that are feasible for many providers
    - Level 2 – requires data interoperability OR measures that are feasible for advanced EHR users
    - Level 3 – integrating patient reported data or other challenging data OR feasible for few, if any, at this time -- stretch goals for the future
    - Other?




Settings of care

EHR adoption in different settings of care is highly variable. Facilities and larger organizations usually have IT resources and infrastructure that small clinician office practices might not have.

Discussion question:

- Does the setting (ambulatory, inpatient, ED, long term care) influence eMeasure feasibility?

Comparability of feasibility and assessment approaches in different EHR systems

The environmental scan of current approaches to assess feasibility is provided.

Discussion questions:

- What are the differences and similarities in approach to assessment of feasibility?
- Do EHR vendors approach assessment of feasibility differently depending on the focus of their services (specialty vendors; large vs. small; setting specific)?
- Is there a different approach to assessing feasibility for registries and EHR systems?
- How comparable are the various approaches in determining feasibility?

Relationship of Feasibility assessment to Reliability and Validity Assessment

NQF has provided some guidance around assessment of eMeasures for reliability and validity as part of the 2010 Measure Testing Task Force Guidance Report (see Appendix for excerpt regarding EHRs). The guidance describes the criteria that NQF Steering Committees should use to evaluate both de novo and retooled eMeasures.

The Measure Testing Task Force Report highlights challenges for eMeasures:

> *Despite these potential advantages over current data sources, several potential sources of error pose threats to the reliability and validity of data elements and computed measure scores for EHR measures including: 1) incorrect measure specifications, including code lists, logic, or computer readable programming language; 2) EHR system structure or programming that does not comply with standards for data fields, coding, or exporting data; 3) difference in use of data fields by different users or entry into the wrong EHR field; 4) entry of incorrect information; and 5) incorrect parsing of data by natural language processing software used to analyze information from text fields. All of these potential errors are analogous to sources of error with measures based on other data sources.*

Discussion questions:

- Are the potential sources of error listed in the Task Force Report than should be addressed through feasibility assessment?
- Is an assessment that demonstrates sufficient feasibility a prerequisite for reliability and validity of the data elements? Can an eMeasure that is not feasible be valid?
- Is feasibility assessment in real world clinical settings distinct from reliability and validity assessment of the data elements?
- How should information from feasibility assessment be incorporated into reliability and validity testing (e.g., the effect of systematic missing data on validity?)
- When should feasibility assessment be performed during the measure development cycle?

- Should there be differences in feasibility assessment for de novo measures and retooled measures as for reliability and validity assessment?

Assessment of Clinical Workflow as part of Feasibility Assessment

Stakeholders support including implementation and assessment of clinical workflow issues as part of feasibility assessment and suggest the following be addressed during feasibility assessment:

- What is the effect of clinician use of EHRs on data element feasibility? Can required information on data elements be recorded correctly in a structured format?
- Can the data elements consistently obtained during clinical practice?
- Does the measure use data elements that are not typically included in structured data fields?

Discussion questions:

- What are realistic expectations for clinician learning and adaptation to structured data within EHR environments?
- What are realistic expectations for adjustments in workflow that would make a measure feasible?
- When are the workflow disruptions sufficiently large to suggest that a measure is not feasible?
- What other implementation challenges diminish eMeasure feasibility?

**RECOMMENDATIONS FOR eMEASURE FEASIBILITY ASSESSMENT**

The guidance document from this project will contain recommendations from this Panel for feasibility assessment to provide a common set of expectation for assessing feasibility of eMeasures. Standardization of eMeasure feasibility assessment will facilitate eMeasure development and implementation in the field.

The Panel is asked to consider the following questions in formulating a set of recommendations for eMeasure feasibility assessment:

- Should both vendor product assessment and real world assessment be required? Is one preferred?
- Who is responsible for feasibility assessment?
- Most stakeholder support assessment in multiple vendor systems. What is the minimum number of systems? What is the ideal number of systems?
- Should eMeasures feasibility be assessed in multiple sites or settings?
- Should there be a minimum number of patients/records tested at each site?
- What are the minimum expectations for feasibility assessment?

- What assessment of implementation and workflow issues is expected?
- Should there be standardized feasibility tests that would allow for comparisons?

**DRAFT STARTER SET OF CRITERIA FOR eMEASURE FEASIBILITY ASSESSMENT**

Steering Committees that evaluate measures for endorsement by NQF use a standard set of evaluation criteria. The NQF measure evaluation criterion for feasibility has three sub-criteria:

- a. For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).
- b. The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.
- c. 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 assessment demonstrates that it is ready to put into operational use).

Discussion questions:

- What specifics for each sub-criteria of Feasibility should be called out for eMeasures?
- Should Steering Committees that are evaluating eMeasures for NQF endorsement expect the following information from feasibility assessment to determine whether an eMeasure meets the Feasibility criteria:
    i. All data elements in structured data fields or can get to structured data (NLP, image, entry by secondary source)
        - Defined fields
        - Standardized value sets
        - Consistent with QDM and MAT
    ii. Measure specifications and logic appropriately embedded by the vendor
    iii. Measure score generated automatically
    iv. Feasibility assessment results from multiple EHR systems
    v. Evaluation of workflow issues
        - Missing data
        - Is the information recorded in the correct fields?
        - Is important information being recorded in an unstructured format?
        - Other implementation assessments
    vi. Conclusions on feasibility of eMeasure
        - Technical problem encountered during feasibilityassessment
        - What actions were taken if initial assessment results were suboptimal?
        - Any specific implementation notes, specifications or directives
- Are there other aspects of feasibility that should be required from assessment?

APPENDIX Excerpt from

# Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties

## Electronic Health Records and Electronic Measures

Development and implementation of electronic health record (EHR) systems hold great promise for the efficient collection of clinical data that can be used for quality measurement. National initiatives call for the adoption of EHRs that include the capability for quality measurement, and NQF has made endorsing quality measures specified for EHRs an important goal. Data stored in EHRs facilitate reporting of quality measures because EHR data 1) are clinically specific, 2) include a large variety of data types including physiologic data such as laboratory values, and 3) decrease the burden of data collection through automated identification, extraction, computation, and aggregation.

Although the concepts of reliability and validity apply equally to measures derived from EHRs, the EHR presents additional issues related to measure testing. Widespread EHR data are not yet available for measure development and testing. In addition, because there are numerous EHR vendors and home-grown EHR systems, it can be difficult to insure that the selected data fields of interest for any particular measure are comparable among different EHRs. Recommendations regarding testing and evaluation of EHR measures are addressed in Section III.

## III. Recommendations for Measures Specified for EHRs

The EHR holds significant promise for improving the measurement of healthcare quality. The availability of a broad range of reliable and valid data elements for quality measurement without the burden of data collection is widely anticipated. Because clinical data can be entered directly into standardized computer readable fields, the EHR will be considered the authoritative source of clinical information. Quality measures based on EHRs use clinical information recorded by healthcare clinicians in discrete computer readable fields; therefore, measurement errors due to manual abstraction, coding by persons other than the originator, or transcription could be eliminated. Despite these potential advantages over current data sources, several potential sources of error pose threats to the reliability and validity of data elements and computed measure scores for EHR measures including: 1) incorrect measure specifications, including code

lists, logic, or computer readable programming language; 2) EHR system structure or programming that does not comply with standards for data fields, coding, or exporting data; 3) difference in use of data fields by different users or entry into the wrong EHR field; 4) entry of incorrect information; and 5) incorrect parsing of data by natural language processing software used to analyze information from text fields. All of these potential errors are analogous to sources of error with measures based on other data sources.

Table 4 provides the guidance for rating the level of evidence for reliability and validity of EHR measures, and it is analogous to the ratings in Table 2 [for traditional measures]. Table 3 indicates how the ratings are used to make a determination if the *Scientific Acceptability of Measure Properties* criterion has been met for EHR measures. Approaches to testing the reliability and validity of the EHR measure score are the same as for any measure as noted in Tables A-1 and A-3.

Tables 2 and 4 differ in two ways. First, EHR measures must be specified in accordance with the Quality Data Model (QDM, formerly called the QDS).[14] The reason for requiring specifications using the QDM is twofold: 1) the QDM can be translated to computer-readable specifications that can be applied to EHRs; and 2) the structure of the QDM will help fulfill the criterion for precise specifications. The QDM will be updated on a regular basis; therefore, if a measure needs a quality data element that is not currently available, then there will be a process to consider additional quality data elements so that the measure could achieve a moderate or high rating.

Second, data elements for quality measures, which are extracted from EHRs using computer programming, are by virtue of automation repeatable (reliable); however, they can be wrong (invalid). Different uses of an EHR data field by clinicians or different data processing or extraction protocols in different EHRs can result in incorrect or missing data and produce different performance scores. Therefore, testing at the data element level should focus on validity as discussed below. Focusing on validity testing of data elements is consistent with the rating

system for all measures presented in Table 2—that is, if empirical validity testing of the data elements is conducted, then separate reliability testing of the data elements is not required.

An approach to testing the validity of data elements analyzes the agreement between data elements and scores obtained with data exported electronically using the EHR measure specifications to those obtained by review and abstraction of the *entire* EHR, preferably using EHRs that comply with standards. This approach has been reported in the literature[15-17] and by HealthPartners in a Commonwealth Fund report[18] on performance measures and EHRs. As with measures for other data types, testing may be conducted on a sample of the measured entities (see Section I).

Because EHR databases may not be available for such testing, another approach is to apply the EHR measure to a simulated data set that reflects standards for EHRs and includes sample patient data with the elements needed for the specified measure. Because the simulated data set is constructed, the values for the data elements and scores are known. When the EHR specifications are applied to the simulated data set, they should return the known values of the data elements and scores.

With either approach, when the results obtained for the EHR measure do not match the known values in the simulated data set or the abstracted data, an analysis is conducted to determine the source of error. If the error is related to the measure specifications, including code lists, logic, and computer readable programming language, then it would be corrected before submission for endorsement. If the source of error is due to clinical data entry practices and EHR structures unique to specific organizations, then the error would not be mitigated by changes to the EHR measure specifications, but it could indicate the need for further evaluation of feasibility and for alternative data fields.

The recommended approach to evaluating reliability and validity of data elements for EHR measures accounts for the current environment in which standards for EHRs and EHR measures

are under development and have not yet been widely adopted. Therefore, testing sites are limited, and testing in a sample of EHR systems may not be representative of all systems. However, this is no different from testing the data elements for measures based on other data sources in a sample of the measured entities whose data practices may vary. As noted in the Background, reliability and validity are not static properties, and no one test is definitive.

Measure testing requirements should not impede the adoption of EHRs and EHR measures, but they should be true to the principles of scientific acceptability of measure properties. EHRs and EHR measures are new and will most likely require some adjustment of local EHR structures and recording practices to meet standards. Therefore, providers should be encouraged to conduct their own internal reliability studies.

Previously endorsed measures specified for chart abstraction or administrative claims data may be appropriate for re-specification for EHRs. Although these endorsed measures should have already been tested for reliability and validity, the EHR measure specifications must be assessed for similarity to the original specifications, which also is addressed in Table 4. In some cases, the EHR specifications will represent a substantive change to the measure so that an assessment of reliability and validity of the EHR measure also is needed.

**Table 4: Evaluation of Reliability and Validity of Measures Specified for EHRs**

| | New Measure Specified for EHR | | Modifications for Endorsed Measure *Re-specified* for EHRs |
|---|---|---|---|
| **Rating** | **Reliability Description and Evidence** | **Validity Description and Evidence** | |
| **High** | All EHR measure specifications are unambiguous[+] and include only data elements from the Quality Data Model (QDM)***** including quality data elements, code lists, and measure logic; **OR** new data elements are submitted for inclusion in the QDM;<br>**AND**<br>Empirical evidence of reliability of both data element **AND** measure score within | The measure specifications (numerator, denominator, exclusions, risk factors) reflect the quality of care problem (1a,1b) and evidence cited in support of the measure focus (1c) under *Importance to Measure and Report*;<br>**AND**<br>Empirical evidence of validity of both data elements **AND** measure score within acceptable norms:<br>• Data element: validity demonstrated by analysis of agreement between data elements electronically extracted and data elements visually abstracted from the entire EHR with statistical results within acceptable norms; **OR** | The EHR measure specifications use only data elements from the Quality Data Model (QDM)***** and include quality data elements, code lists, and measure logic;<br>**AND**<br>Crosswalk of the EHR measure specifications (QDM quality data elements, code lists, and measure logic) to the endorsed measure specifications demonstrates that they represent the original measure, which was judged to be a valid indicator of quality; |

APPENDIX Excerpt from

# Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties

| | New Measure Specified for EHR | | Modifications for Endorsed Measure *Re-specified* for EHRs |
|---|---|---|---|
| **Rating** | **Reliability Description and Evidence** | **Validity Description and Evidence** | |
| | acceptable norms:<br>• Data element: reliability (repeatability) assured with computer programming—**must test data element validity**<br>**AND**<br>• Measure score: appropriate method, scope, and reliability statistic within acceptable norms | complete agreement between data elements and computed measure scores obtained by applying the EHR measure specifications to a simulated test EHR data set with known values for the critical data elements;<br>**AND**<br>• Measure score: appropriate method, scope, and validity testing result within acceptable norms;<br>**AND**<br>Identified threats to validity (lack of risk adjustment/stratification, multiple data types/methods, systematic missing or "incorrect" data) are empirically assessed and adequately addressed so that results are not biased | **AND**<br>Analysis of comparability of scores produced by the retooled EHR measure specifications with scores produced by the original measure specifications demonstrated similarity within tolerable error limits |
| **Moderate** | All EHR measure specifications are unambiguous+ and include only data elements from the QDM;* **OR** new data elements are submitted for inclusion in the QDM;<br>**AND**<br>Empirical evidence of reliability within acceptable norms for either data elements **OR** measure score as noted above | The measure specifications reflect the evidence cited under *Importance to Measure and Report* as noted above;<br>**AND**<br>Empirical evidence of validity within acceptable norms for either data elements **OR** measure score as noted above; **OR**<br>Systematic assessment of face validity of measure score as a quality indicator (as described in Table A-3) explicitly addressed and found substantial agreement that ***the scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality***<br>**AND**<br>Identified threats to validity noted above are empirically assessed and adequately addressed so that results are not biased | The EHR measure specifications use only data elements from the QDM as noted above<br>**AND**<br>Crosswalk of the EHR measure specifications as noted above demonstrates that they represent the original measure<br>**AND**<br>For measures with time-limited status, testing of the original measure and evidence ratings of moderate for reliability and validity as described in Table 2. |
| **Low** | One or more EHR measure specifications are ambiguous+ or do not use data elements from the QDM*;<br>**OR**<br>Empirical evidence of unreliability for either data elements **OR** measure score—i.e., statistical results outside of acceptable norms | The EHR measure specifications do not reflect the evidence cited under *Importance to Measure and Report* as noted above;<br>**OR**<br>Empirical evidence (using appropriate method and scope) of invalidity for either data elements **OR** measure score— i.e., statistical results outside of acceptable norms<br>**OR**<br>Identified threats to validity noted above are empirically assessed and determined to bias results | The EHR measure specifications do not use only data elements from the QDM;<br>**OR**<br>Crosswalk of the EHR measure specifications as noted above identifies that they do not represent the original measure<br>**OR**<br>For measures with time-limited status, empirical evidence of low reliability or validity for original time-limited measure |
| **Insufficient evidence** | Inappropriate method or scope of reliability testing | Inappropriate method or scope of validity testing (including inadequate assessment of face validity as noted above)<br>**OR**<br>Threats to validity as noted above are likely and are NOT empirically assessed | Crosswalk of the EHR measure specifications as noted above was not completed<br>OR<br>For measures with time-limited status, inappropriate method or scope of reliability or validity testing for original time-limited measure |

APPENDIX Excerpt from

# Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties

[+]Specifications are considered unambiguous if they are likely to consistently identify who is included and excluded from the target population and the process, condition, event, or outcome being measured; how to compute the score, etc.

*QDM (formerly called the QDS) elements should be used when available. When quality data elements are needed but are not yet available in the QDM, they will be considered for addition to the QDM.