NATIONAL QUALITY FORUM

Memo

TO: Consensus Standards Approval Committee (CSAC)

FR: NQF Measure Maintenance Team

RE: eCQM Approval for Trial Use Evaluation Plan

DA: December 12, 2017

Background

In 2014, NQF piloted the electronic clinical quality measures (eCQM or eMeasure) Approval for Trial Use (ATU) program to support eCQMs that are ready for implementation, but cannot yet be adequately tested to meet NQF endorsement criteria. The program uses the NQF multistakeholder consensus process to evaluate and approve eCQMs. All eCQMs in this program must meet NQF endorsement criteria to be considered, specifically they must be evidence-based and address important areas for performance measurement and quality improvement. And, the eCQMs must be assessed to be technically acceptable for implementation so that developers may collect data during the three year "trial use" period to support the NQF reliability and validity testing criteria for endorsement consideration.

This memo provides an update on the eCQM Approval for Trial Use program and preliminary evaluation metrics. The goal of the on-going evaluation is to determine if the program is structured in such a way to support development and testing of eCQMs. NQF staff is interested to learn how effective this program is, specifically how it has potentially supported development or testing of eCQMs since its inception.

Program Status

Eleven measures from five developers have been Approved for Trial Use since 2014 including, 9 process measures, 1 intermediate clinical outcome measure and 1 composite measure.

NQF Number	Title	Type	Developer
2522	Rheumatoid Arthritis: Tuberculosis Screening	Process	American College of Rheumatology
2525	Rheumatoid Arthritis: Disease Modifying Anti-Rheumatic Drug (DMARD) Therapy	Process	American College of Rheumatology
2549	Gout: Serum Urate Target	Process	American College of Rheumatology
2550	Gout: ULT Therapy	Process	American College of Rheumatology

NQF Number	Title	Type	Developer
2597	v 1.0: Substance Use Screening and Intervention Composite (0028, 2152, 2596)	Composite	American Society of Addiction Medicine
2721	Screening for Reduced Visual Acuity and Referral in Children	Process	Centers for Medicare & Medicaid Services
2872	Dementia- Cognitive Assessment	Process	PCPI
2983	Potassium Sample Hemolysis in the Emergency Department	Process	Cleveland Clinic
3059	One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk	Process	PCPI
3060	Annual Hepatitis C Virus (HCV) Screening for Patients who are Active Injection Drug Users	Process	PCPI
3061	Appropriate Screening Follow-up for Patients Identified with Hepatitis C Virus (HCV) Infection	Process	PCPI

Initial Feasibility Assessment & Workflow Evaluation Themes

Baseline data are collected to determine if there is uptake in the number of electronic health record (EHR) systems used for feasibility assessment, as well as the number of EHR systems available for testing scientific acceptability when a measure is submitted for initial ATU and for endorsement. To collect this information, NQF staff distributed surveys to five measure developers and received a response rate of 60%. The three respondents represented nine out of the 11 measures in the program.

- Feasibility scores across the measure participating in the program demonstrated moderate or high feasibility based on survey of four EHR vendors. Typically, one to two data elements were unable to be captured in structured fields.

Testing Status & Endorsement Status

- Field testing has commenced on 8 of the measures, in at least 1 – 3 sites.
- Identified EHRs in test sites: Sunquest, EPIC, GE Centricity & Cerner/LabCorp

Key Themes on Developer Survey Responses

Challenges

- Recruitment for testing was challenging particularly because sites are largely not using validated tools and/or standardized methods of EHR reporting for screening or interventions.

- Identifying test sites that are currently collecting all of the required data elements was also challenging.
- Test sites reported that they do not routinely capture medical exceptions and limited life expectancy data, making it difficult to derive a workflow score.
- Not all developers submitted responses to the ATU baseline evaluation survey.

Implementation

No feedback received on implementation.

Benefit of ATU Status

- Allows NQF formal review of concept and supporting evidence while allowing time to test [measure].
- Serves as formal measure validation and gives measures credibility for CMS and other non-quality payment programs that might otherwise not be used.
- The approval for trial use designation can give the measure prominence when applying for grants and funding. At the time of the survey, however, a developer noted that they have been unable to secure funding to complete the testing of their measure.

Discussion Question for CSAC

- Suggestions for further data gathering for evaluation of the program?

Evaluation Plan for NQF Approval for Trial Use Program

Overview

In 2014, NQF piloted the eMeasure Approval for Trial Use program to support eMeasures that are ready for implementation, but cannot yet be adequately tested to meet NQF endorsement criteria. The program uses the NQF multistakeholder consensus process to evaluate and approve eMeasures. All eMeasures in this program must meet basic NQF endorsement criteria to be considered. Specifically they must demonstrate that they address important areas for performance measurement and quality improvement. Also, the eMeasures must be assessed to be technically acceptable for implementation so that data may be collected during the three year "trial use" period to support the NQF endorsement reliability and validity testing criteria.

The goal of the Approval for Trial Use (ATU) program evaluation is to determine if the program is structured in such a way to further support eMeasures along the continuum from development and testing to successful submission for NQF endorsement. The NQF Maintenance Team will evaluate the ATU program performance using a series of metrics and research questions to determine the success and value of the program.

Key Metrics

1. Number of eMeasures submitted to the program v submitted for endorsement
2. Number of eMeasures recommended, accepted and approved in the program by type (e.g. de novo, respecified, etc.)
3. Status of eMeasures in program at the three-year approval period (e.g. endorsed, dismissed, etc.)
4. Number of eMeasures "converted" to NQF Endorsed Status
5. Lessons learned (internal processes to support program)

Evaluation Questions

Primary

1. Should NQF continue to offer the eMeasure Trial Use program? If so, are there any changes that might enhance the program?

Secondary

1. Types of Measures: Are there common types (structure, process, outcome, PRO-PM) of eMeasures that benefit from this program?
    a. Are there specific situations that drive participation? (e.g. funding streams for development, level of analysis of measures, intended use)
    b. Are there other themes or commonalities of the types of measures in the ATU program?
2. Does the ATU program increase the number of EHR sites using eMeasures?
    a. Does ATU increase implementation and use of eMeasures?

b. Are there other themes or considerations of how eMeasures in the program being used?

3. How does the ATU program support endorsement and implementation progress of eMeasures?
    a. Is there a change in eMeasure submission for NQF endorsement as a result of the ATU program?
    b. What is the status of eMeasures in the program as it pertains to readiness for endorsement? (1 year, 2 years, 3 years)
4. Are there lessons learned, or additional information received by the measure developer, relevant to this evaluation? (Barriers, challenges, success?)

Exhibit 1: Research Questions and Identification of Data Sources

	Relevant Data Sources				
Research Questions	Review of Measure Submissions	Measure Application Partnership Review	Survey of Measure Developers	Structured Interviews with Developers/Stewards	Measure Endorsement
What types of measures (including situations that lead to a submission) are being submitted to the Approval for Trial Use program?	√		√	√	√
Does the ATU program use and implementation of EHR sites using eMeasures?			√	√	√
How does the Approval for Trial Use program support the level of endorsement and implementation progress of eMeasures?	√	√	√	√	√
Lessons Learned			√	√	

We will evaluate the initial measure submission forms to gain an initial understanding of each application and the measure developer's plans to transition the eMeasure from the ATU program to be considered for NQF endorsement. This will be supplemented with additional qualitative data gathered through structured interviews and surveys of measure developers to provide additional and practical understanding of the effectiveness of the ATU program.

Evaluation Approach

Our approach is to design a rigorous methodology that provides a valid and comprehensive assessment of individual eMeasures submitted into the ATU program, as well as aggregate assessments to support the evaluation of the program. In Exhibit 2 we provide a proposed model to guide our evaluation approach divided by three phases.

Exhibit 2: Proposed Evaluation Model: Evaluation Components and Concepts

Baseline (at approval for ATU)	**Intermediate** (Year 1, 2 and 3 of ATU period)	**Distal Outcomes** (3 – 5 years post ATU period)

• Types of measures submitted to program
  o Measure type
  o Level of analysis
  o Care setting
• Types of measure gaps filled by ATU measures
• Feasibility assessment scores
• Identification of key stakeholders interested in the outcome of the measure (providers, consumers, public health, plans, long term care, large employers)
• Testing plan
	• Test site engagement and implementation
	• Initial measure scores
	• Workflow evaluation
	• Usability (including organizational readiness)
	• EHR types included in implementation (e.g. integration of measure into more than one EHR)
	• Status of moving measure out of ATU program
	• Final testing results
		• Provider satisfaction with measure
		• Clinical usability
		• Patient engagement with measure (if applicable)
		• Sustainability of measure if approved for endorsement

At the far left side, we list a series of inputs that correspond to a baseline evaluation of each measure submitted to the ATU program. These starting conditions contain key metrics such as the type of measure being submitted; the type of measure gap being fulfilled; the results of the initial feasibility assessment; the stakeholders involved in the development and use of the measure; and the types of data needed for testing. As we evaluate these measures after acceptance into the program, we examine inputs that build off of the initial information and begin to formulate a summative (intermediate) evaluation to understand if a measure in the program was successfully tested and used to meet the objectives to be considered for potential endorsement.

Three years post ATU period, we may look at expected program outcomes for a summative evaluation, including provider satisfaction with the measure, sustainability and impact of the

measure. However, due to the initiation of the *Graduated Measure* program and the proposed integration of ATU, we will not include data collection plans for this phase.

Data Collection Methods

Details on measure use and testing can be collected through a review of the measure submission form information in OPUS, final committee reports, measure developers surveys and interviews, monitoring the work of Measure Application Partnership (MAP) workgroups, , and tracking annual updated and other follow up measure endorsement activities. Data collection will also include detail about the individual measure developer and classification of measure topic and type, as this information will contribute to a better understanding of program use and performance. To facilitate collection of information that will inform program evaluation, measure developers should be prepared to communicate with NQF during the Approval for Trial use program period on an annual basis.

Data collected can be leveraged for the evaluation using the following methods:

Review of Measure Information sourced from NQF systems:

NQF will examine measure information from a series of internal NQF systems including the initial measure submission information; the feasibility scorecard; the initial testing results from BONNIE or another test data extract; the importance to measure and report; initial discussions on reliability and validity; and the type of measure being submitted and what gap(s) it could potentially fulfill. This information will be culled from OPUS, Committee Reports and SharePoint project sites.

Measure Application Partnership:

NQF Staff will also track measures proposed for use in federal programs through the final reports of each of the MAP workgroups. NQF will identify measures that have received Approval for Trial Use from the Measures Under Consideration (MUC) inventory provided yearly by the Department of Health and Human Services (HHS), and report on the result of MAP workgroup deliberations on these measures. In addition, any measures on the MUC list that come into the ATU program will also be tracked.

Measure Developer Survey:

The measure developer should provide information regarding measure use in the field and measure testing details from measure developers. This information should be collected through a survey, beginning 12 months after final approval of Trial Use status and on a yearly basis associated with the measure annual update during the program period.

Initial Questions for Measure Developers:

- How did developer find out about the program? (Did they submit for NQF endorsement, or had they been referred to the program by NQF as a result of technical assistance?).
- What were the barriers or difficulties in testing the measure that necessitated submission to the Approval for Trial Use program?
- Who are the key stakeholders for the measure (development and implementation)?
- Do you have a testing plan?
- Who will be responsible for collecting data on the measure?

Annual Update Survey Questions:

- How many sites is the measure being tested at? And how many EHRs has the measure been implemented on? If yes, please provide names of sites and vendors.
- How many providers are using the measure?
- Based on preliminary testing results, have any refinements been made to the measure?
- Please describe how the measure has been integrated into clinical workflow.
- Does the measure developer anticipate changes to the measure specifications approved by NQF? If yes, please describe.
- Do you anticipate submitting the measure to be considered for NQF endorsement this year?

Structured Interviews with Measure Developers: We will conduct interviews with measure developers on an annual basis as a follow-up to information gathered in the annual update survey. The goal of these interviews will be to gather qualitative feedback on challenges, barriers with testing of the measure and the ATU program.

Measure Endorsement: An important consideration for evaluation of the ATU program is the number of final submissions for endorsement during and immediately after the three-year approval period, and the timing and frequency of measure submissions for endorsement relative to approval date. Additional considerations include the results of Standing Committee deliberations during measure review, and significant changes in measure specifications or intent from the original ATU submission.