NATIONAL QUALITY FORUM

Measure Developer Workshop 2019

June 10, 2019

Agenda

- **Welcome and Introductions**
- **The NQF Process and Submission Guidance**
  - *Pulse Check: NQF Submission Process*
  - *NQF's Methods Panel – What Good Looks Like*
  - ***Break***
  - *Approval for Trial Use Program: Opportunity to Access NQF Services for Measures in Development*
  - ***Lunch***
- **Emerging Topics in Measurement Science**
  - *Social Risk Trial: Data Collection and CDP Submissions*
  - *Measure Sets and Measurement Systems*
  - ***Break***
  - *Human-Centered Design in Practice for Measurement Development*
  - *Measure Feedback Loop*
  - *eCQM Submission Considerations*
- **Summary of the Day and Next Steps**

The NQF Process and Submission Guidance

NATIONAL QUALITY FORUM

Pulse Check: NQF Submission Process

Karen Johnson, Senior Director, NQF
Jean-Luc Tilly, Senior Data Analytics Manager, NQF

Session Overview

- Background
  - *History of the redesign*
  - *Major elements of the redesign*
- Discussion
  - » Two-cycle structure
  - » Intent to Submit
  - » Methods Panel

History of the Redesign

- Motivation for the redesign
  - *Stakeholder concern about NQF's agility*
    - » Time from measure submission to measure endorsement
    - » Timeliness of measure evaluation/wait time for available projects
- Approach
  - *Kaizen event on May 18-19, 2017, using LEAN tools*
- Participants
  - *>40 attendees + NQF staff/consultants*
  - *Public and private sector stakeholders*
    - » CMS and other federal agencies
    - » NQF standing committee members
    - » Measure developers

History of the Redesign

Objectives

- To examine the timeliness, efficiency, and effectiveness of the CDP, with a view toward identifying its strengths and weaknesses and where it might be improved using a more agile process
  - *Continuous availability of CDP for all measure types*
  - *Improved management of the CDP measure pipeline*
  - *Improved utilization of standing committee expertise*
  - *Improved leveraging of NQF and external expertise*
  - *Reduction in overall endorsement time to about 6 months*

History of the Redesign

Approach

- Stream #1: Measure pipeline and scheduling
  - *Improving coordination among CMS, developers, and NQF to better facilitate timely evaluation of measures*
- Stream #2: Streamlining the CDP
  - *Increasing opportunities for submission and timely review of measures*
  - *Reducing cycle time of the CDP*
- Stream #3: MAP/CDP integration
  - *Improving flow of information between the CDP and Measure Applications Partnership (MAP) processes*

Major Elements of the Redesign

- Scheduling/frequency: Two evaluation cycles per year
  - *Topic area consolidation (from 22 to 15)*
- Intent to Submit process
  - *Meant to help facilitate planning of evaluations*
  - *Required for implementation of the SMP*
- Scientific Methods Panel (SMP)
  - *Reduce standing committee (SC) burden*
  - *Promote consistency in evaluation of reliability and validty*
  - *Encourage greater participation in SCs by consumers, patients, and purchasers*

Main Elements of the Redesign

- Continuous commenting
  - *Consolidated pre- and post-comment periods to allow more time*
  - *Replaced NQF member voting with member expression of support*
- Technical report restructuring
  - *Shorten the report (some content to be moved to website)*
  - *Produce annual cross-cutting report*
- Education
  - *Facilitation training for staff and co-chairs*
  - *On-demand webinars for developers, NQF members, and the public; additional training for staff*
- Data integration for CDP and MAP

Cycle Scheduling

The new cycle schedule resulted in 3 significant changes to the timing of the endorsement process:

- More frequent opportunity to submit measures (2 per year in each topic area)
- All projects aligned to the submission deadline schedule
- Shorter cycle time for evaluation (6 months: submission deadline through endorsement decision)

Intent to Submit

- Original goals
  - *To clarify the evaluation schedule for measures*
  - *Provide a "pipeline" of measures to help NQF better schedule evaluations*
- In practice
  - *also used to facilitate the new SMP process*
  - *Measure specifications and testing attachment required at ITS deadline*

Scientific Methods Panel

- Original goals
  - *Reduce standing committee (SC) burden*
  - *Promote consistency in evaluation of reliability and validity*
  - *Encourage greater participation in SCs by consumers, patients, and purchasers*
- In practice
  - *The process has evolved substantially over the past 4 cycles*
  - *SC burden has decreased, due to "gatekeeper" role*
  - *The SMP's advisory role has provided value to the field*
    - » This will become even more apparent in the next year

Survey Results: Developers

- Total number of responses: 10
- # measures submitted, on average, since redesign:
  - *Ranged from 2 to 25*
- Approximate percentage of measures evaluated by the SMP:
  - *Ranged from 0% to 100%*
- Experience with NQF since redesign
  - *Brand new: 20%*
  - *Not new, but don't interact often: 40%*
  - *Very experienced (submit a lot of measures): 40%*

Survey Results: Developers

- Frequency of 6-month cycle time (n=10)

Too frequent	Not frequent enough	Just right	Not sure
10%	0%	70%	20%

- Has number of simultaneous evaluation projects negatively impacted work (n=10)

Not at all	Somewhat	Substantially	Not sure
40%	50%	0%	10%

Survey Results: Developers

- Does the SMP add value to the CDP (n=9)
  - *Yes: 22%*
  - *No: 78%*
- Some comments
  - *Extra level of review, more requirements, not transparent*
  - *Increased complexity, don't have consensus on what meets requirements, harder to pass R/V*
  - *SMP knowledge of survey-based measures inadequate*
  - *Another hurdle; dislike gatekeeper function*

Discussion: Cycle Scheduling

- How has the cycle scheduling worked for you?
  - *What do you like?*
  - *What don't you like?*
  - *What are the pain points?*
- Should NQF continue with this schedule for all topic areas?
  - *Would changing help or hinder your submission efforts?*
- If we consider a change to the schedule, what do you want us to think about?
- Do you have suggestions for changes?

Discussion: Intent to Submit

- How has the ITS process worked for you?
  - *What do you like?*
  - *What don't you like?*
  - *What are the pain points?*
- What has discouraged you from using the ITS process as a way to communicate your "pipeline" of measures?
- Have you sought technical assistance prior to/during the ITS period?
  - *Why or why not?*
  - *What hasn't gone as well as you'd like?*
- Do you have ideas on how NQF can better help you before/during the ITS period?

Discussion: Scientific Methods Panel

- Do you believe the SMP has added value to the CDP?
  - *Why or why not?*
- What do you think about the changes made to the SMP process over the past four evaluation cycles?
- What are your thoughts about the SMP's "gatekeeper" role?
- Do you have ideas on how NQF could improve the SMP process?

NATIONAL QUALITY FORUM

NQF's Scientific Methods Panel – What Good Looks Like

Karen Johnson, Senior Director, NQF
J. Matt Austin, PhD

Introductions

- Karen Johnson, Senior Director, NQF
- J. Matt Austin, PhD, Assistant Professor, Armstrong Institute for Patient Safety and Quality, Johns Hopkins Medicine (and member of NQF's Scientific Methods Panel)

Background

- Scientific Methods Panel (SMP) implemented as part of NQF's 2017 redesign of the Consensus Development Process (CDP)
  - *Evaluate the scientific acceptability of complex measures*
  - *Advise NQF on methodologic issues*
- Complex measures
  - *Outcomes, instrument-based, cost/resource use, efficiency, and composites*
- To date, the SMP has evaluated 115 measures since the fall 2017 evaluation cycle
  - *In fall 2018/spring 2019 cycles, approximately one-quarter of the SMP's measures did not "pass" reliability or validity*

Specifications

The criterion: The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability

What good looks like:

- Enough detail to understand how to calculate and implement the measure
  - *Beyond detailed numerator, denominator, and exclusions statements, also address things like the sampling methodology (if allowed), use of proxy responses, how missing data are handled, required minimum number of cases, etc.*
  - *Take advantage of "calculation algorithm" item*
  - *May need brief "supplements" to explain special circumstances (e.g., use of a particular data source)*

Testing Methodology

What good looks like:

Provide details!!

- Don't just name an approach/method/statistic – explain what you did
  - *Example:* saying you did a signal-to-noise analysis isn't sufficient, as there are multiple ways to do this
  - *Example:* saying you did data element validation isn't sufficient; instead, describe how you did it (also, be clear about who/what you are using as the gold standard)
  - *Example:* a recent measure needed description from the statistical programming language manual to understand exactly what statistic was being reported (saying "ICC=x" wasn't enough)

Testing Results

What good looks like:

- **Provide details and be organized**
  - Use “headers” to orient readers
  - Be careful with table labels—be precise about what the table is showing
  - Help put results into context: Remind the readers about the testing samples (e.g., include n’s)
  - When appropriate, give more than a one-number result (e.g., mean reliability); also, provide information about the variation (e.g., standard deviation, other distributional statistics, estimates for various sample sizes)
- Interpret results
  - Don’t just say something like “this means the measure is highly reliable”—instead, explain how/why you make this conclusion

Construct Validation

- Provide narrative describing the hypothesized relationships
- Explain why you think comparing these measures would validate the measures
- State the expected direction of the association
- State the expected strength of the association
- Name and describe the specific statistical tests used
- Provide results
- Interpret the results
  - *How do the results "match" your hypotheses?*
    - » If not, address it and discuss why
  - *How do the results help to validate the measure?*

Additional Considerations

- Ensure that testing matches NQF requirements
  - *NOTE that NQF has different requirements depending on measure type and whether new vs. maintenance*
- Ensure that testing matches specifications
  - *Level of analysis, care setting*
  - *Do not test using a minimum threshold, unless it is part of the measure specification*
  - *If risk-adjusted, test the risk-adjusted result (and state that is what you did!)*
- Adjusting for social risk
  - *Conceptual rationale is required, even if measure is not risk-adjusted or if risk-adjustment does not include social risk factors*

Break

NATIONAL QUALITY FORUM

Approval for Trial Use Program:
For Measures in Development

Wunmi Isijola, Senior Managing Director, NQF

What is Approval for Trial Use (ATU)?

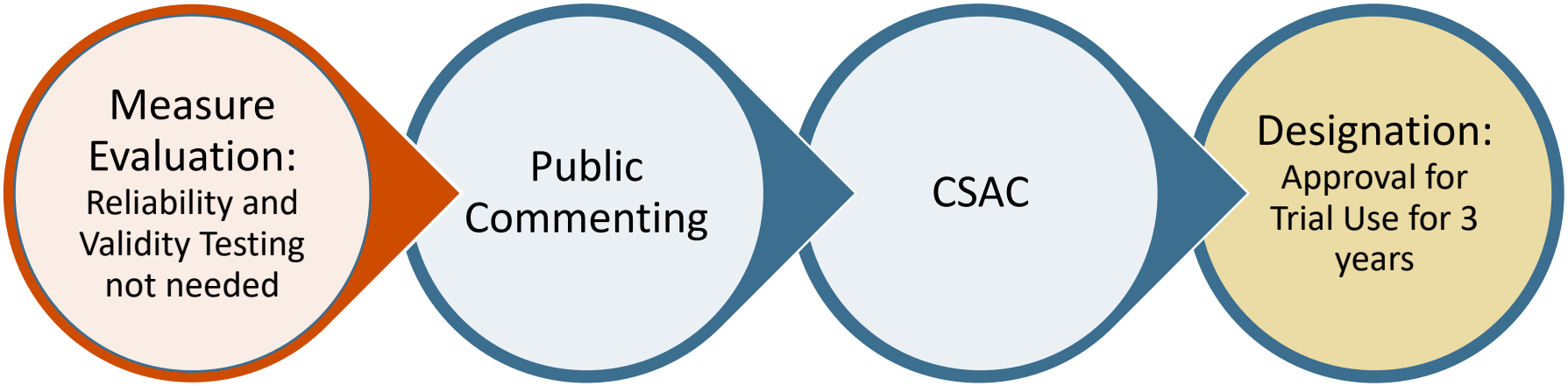
Launched in 2014, The Approval Trial Use Program is intended:

- Primarily for eCQMs that are ready for implementation but cannot be adequately tested to meet NQF endorsement criteria
- To promote implementation and conduct robust reliability and validity testing of clinical data in EHRs

Current Criteria for ATU

- Must meet all criteria under NQF's *Importance to Measure and Report.*
- Complete the eCQM feasibility assessment.
- Results from testing with a simulated (or test) data set.
- Plan for future use and discussion of how the measures will be useful for accountability and improvement.
- Related and competing measures are identified with a plan for harmonization or justification for developing a competing measure.

ATU Process


Measure Evaluation:
Reliability and Validity Testing not needed
Public Commenting
CSAC
Designation:
Approval for Trial Use for 3 years

eCQMs Submitted through ATU

2522 Rheumatoid Arthritis: Tuberculosis Screening
2525 Rheumatoid Arthritis: Disease Modifying Anti-Rheumatic Drug (DMARD) Therapy
2523 Rheumatoid Arthritis: Assessment of Disease Activity
2549 Gout: Serum Urate Target
2550 Gout: ULT Therapy
2597 Substance Use Screening and Intervention Composite
2721 Screening for Reduced Visual Acuity and Referral in Children
2764 Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy
2872 Dementia- Cognitive Assessment
2983 Potassium Sample Hemolysis in the Emergency Department
3059 One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk
3060 Annual Hepatitis C Virus (HCV) Screening for Patients who are Active Injection Drug Users
3061 Appropriate Screening Follow-up for Patients Identified with Hepatitis C Virus (HCV) Infection

Approval for Trial Use (ATU)

The National Quality Forum (NQF) is exploring expanding ATU to all types of measures to better serve the quality measurement community and address changing needs in the field.

Goal of expansion is to:

- Encourage measure innovation and close critical gaps
- Reduce measure development costs and measurement burden

Opportunities under ATU

- Guidance on NQF Processes
- Standing Committee Review
- Public Commenting
- Approval for Trial Use
  - *Approval designation automatically expires 3 years after initial approval if not submitted for endorsement prior to that time.*
  - *If submitted for endorsement prior to the 3-year expiration, the developer can select from the following options for evaluation and endorsement:*
    - » Submit and evaluate only Scientific Acceptability of Measure Properties, including the final eCQM specifications and all testing.
    - » Submit and evaluate on all criteria.

Endorsement vs Trial Approval

Endorsement	Trial Approval
The eCQM has been evaluated to **meet all NQF evaluation criteria and is suitable for use in accountability applications** as well as performance improvement.	The eCQM has been evaluated to **meet the criteria that indicate its readiness for implementation** in real-world settings in order to generate the data required to assess reliability and validity.

Questions?

Lunch

Emerging Topics in Measurement Science

NATIONAL QUALITY FORUM

Social Risk Trial: Data Collection and CDP Submissions

Nicolette Mehas, Director, NQF
Erin O'Rourke, Senior Director, NQF
Measure Developer Co-presenter

Background

What is risk adjustment?

Risk adjustment is a statistical approach that allows patient-related factors (e.g., comorbidity and illness severity) to be taken into account when computing performance measure scores, thereby improving the ability to make fair and correct conclusions about quality. Although there are various ways to risk adjust, the most common method is use of multivariable statistical models.

Why risk adjust?

- Patients are not randomly assigned to healthcare units, and the characteristics of the patients treated varies across healthcare unit
- Avoid incorrect inferences
- In the context of comparative performance assessment, the general question being addressed is:
  - *How would the performance of measured entities compare if, hypothetically, they had the same mix of patients?*

Context

- In 2014, NQF convened an Expert Panel to review the NQF policy prohibiting the inclusion of social risk factors.
  - *The Panel recommended allowing the inclusion of social risk factors when there was a conceptual and empirical basis*
- NQF Board approved a two-year trial period when social risk factors could be included
  - *The first trial demonstrated that adjusting measures for social risk factors is feasible but challenging*
- NQF has launched a new three-year initiative to continue examining the impact of social risk factors

Social Risk Adjustment: Expert Panel Guidance

- Each measure must be assessed individually to determine if social risk adjustment is appropriate.
- Not all measures should be adjusted for SDS.
  - *Need conceptual basis (logical rationale, theory) and empirical evidence.*
- Recommendations apply to any level of analysis including health plans, facilities, and individual clinicians.
- During the trial period, if adjustment was determined to be appropriate for a given measure, NQF endorses one measure with specifications to compute:
  - *Adjusted measure with social risk factor(s)*
  - *Nonadjusted version of the measure (clinically adjusted only) to allow for stratification of the measure by social risk factor(s)*

Key Recommendations from the Risk-Adjustment Expert Panel

- **Recommendation 1**: When there is a conceptual relationship between sociodemographic factors and outcomes or processes of care and empirical evidence that sociodemographic factors affect an outcome or process of care reflected in a performance measure:
  - *those sociodemographic factors should be included in risk adjustment of the performance score (using accepted guidelines for selecting risk factors) unless there are conceptual reasons or empirical evidence indicating that adjustment is unnecessary or inappropriate; AND*
  - *the performance measure specifications must also include specifications for stratification of a clinically-adjusted version of the measure based on the sociodemographic factors used in risk adjustment.*

- **Recommendation 4:** The NQF criteria for endorsing performance measures used in **accountability applications** (e.g., public reporting, pay-for-performance) should be revised

Key Recommendations from the Risk-Adjustment Expert Panel Continued

- **Recommendation 6:** When there is a conceptual relationship and evidence that sociodemographic factors affect an outcome or process of care reflected in a performance measure submitted to NQF for endorsement, the following information should be included in the submission:
  - *Rationale and decisions for selecting or not selecting sociodemographic risk factors and methods of adjustment*
  - *In addition to identifying current and planned use of the performance measure, a discussion of the limitations and risks for misuse of the specified performance measure.*

- **Recommendation 7:** NQF should consider expanding its role to include guidance on implementation of performance measures. Possibilities to explore include:
  - *guidance for each measure as part of the endorsement process;*
  - *guidance for different accountability applications*

Key Recommendations from the Risk Adjustment Expert Panel Continued

- **Recommendation 8:** NQF should make explicit the existing policy that endorsement is for a specific context as specified and tested for a specific patient population, data source, care setting, and level of analysis. Endorsement should not be extended to expanded specifications without review and usually additional testing.

- **Recommendation 9:** When performance measures are used for accountability applications users of performance measures should assess the potential impact on disadvantaged patient populations and the providers/health plans serving them to identify unintended consequences and to ensure alignment with program and policy goals. Additional actions such as creating peer groups for comparison purposes could be applied.

- **Recommendation 10:** NQF should develop strategies to identify a standard set of sociodemographic variables (patient and community-level) to be collected and made available for performance measurement and identifying disparities

NQF Social Risk Trial Overview

- In May 2018, NQF was awarded a contract from CMS to implement a Social Risk Factor Trial to review measures submitted for initial endorsement or maintenance review over a three-year period.
- NQF and the Disparities Standing Committee will:
  - *explore the inclusion of social risk factors in risk-adjustment models, and*
  - *inform NQF's policy on whether or not to allow the inclusion of such factors in measures submitted for endorsement.*

Project Approach and Scope

In order to meet the project goals, NQF will:

- allow measure developers to submit measures for endorsement with social risk factors included in their risk-adjustment model,
- explore unresolved issues from the initial trial period to advance the science of risk adjustment, and
- explore the challenges and opportunities related to including social risk factors in risk-adjustment models.

Activities to Date

- NQF has reviewed and compiled measure information for measures that were submitted to the fall 2017 through spring 2019 endorsement review cycles.
  - *Located on the Social Risk Trial project page; updated every 6 months.*
  - *Also available is a FAQ guidance document for members of the public who may wish to follow or engage with the measures included in the trial.*
- NQF has met twice with the Disparities Standing Committee.

Overview of Spring 2019 Cycle Submissions

93 measures submitted
- 45 measures assessed outcomes (includes intermediate outcomes and PRO-PMs)

30 measures utilized some form of risk adjustment
- 28 measures provided a conceptual rationale for potential impact of social risk factors
  - 19 measures used literature to support, 15 measures used data (not mutually exclusive)

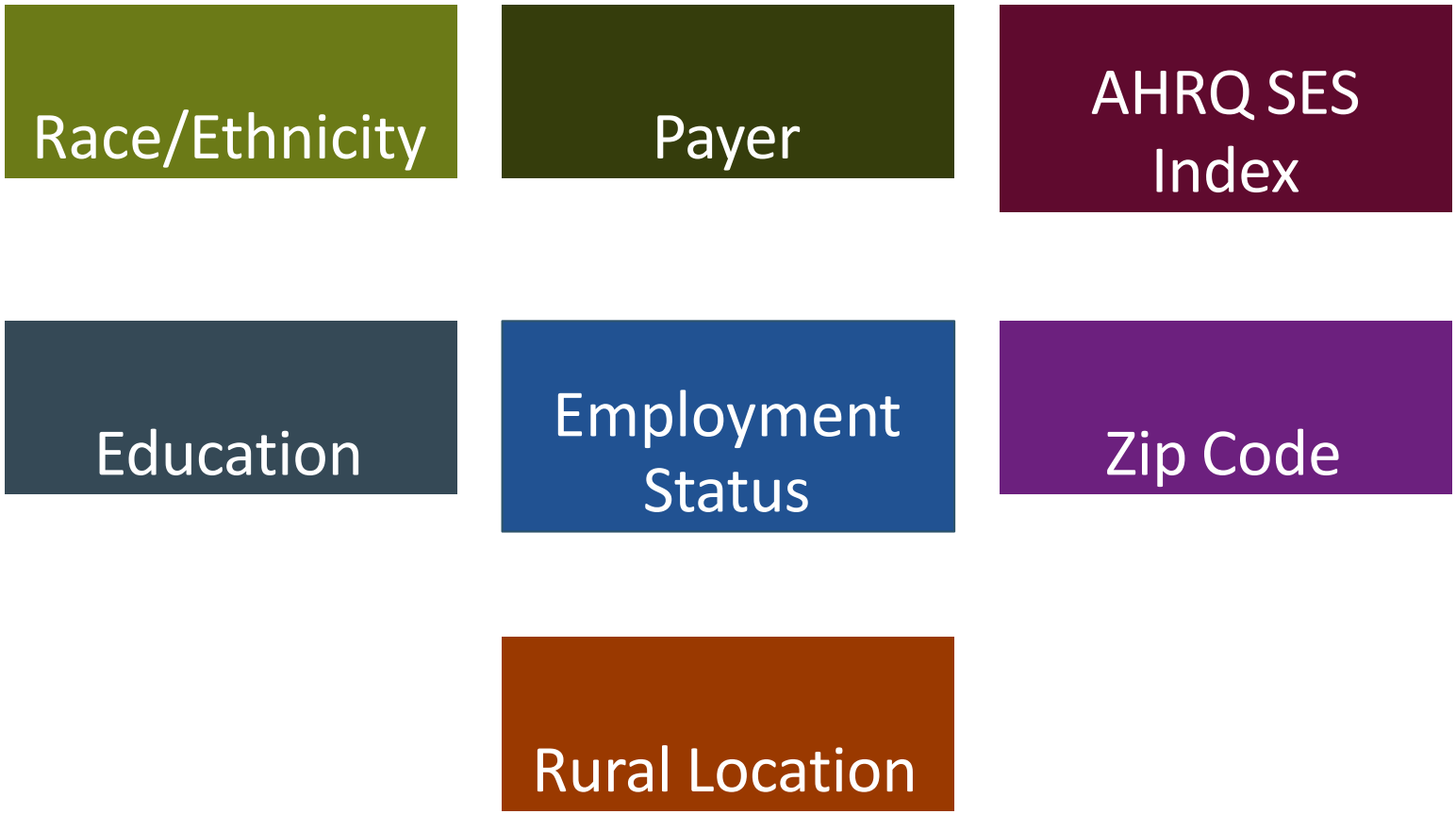
21 measures – conceptual rationale supported the inclusion of social risk factors
- 14 measures – limited or no impact on model performance; social risk factors not included
- 7 measures adjusted for social risk factors

Summary of Submissions: Fall 2017 - Spring 2019

Total Number of Measures Submitted	237
Measures Using Risk Adjustment	100
Measures with a Conceptual Model Outlining Impact of Social Risk*	91
Used published literature to develop rationale	69
Used "Expert Group Consensus" to develop rationale	13
Used "Internal Data Analysis" to develop rationale	47
Measures with a Social Risk Factor included in Model	25

*methods not mutually exclusive

Common Social Risk Factors Considered Fall 2017 - Spring 2019



Early Findings

- Many developers continue to examine race as a potential variable.
  - *However, some do not consider it a social risk factor*
- Disconnect between conceptual relationship and empirical analysis
  - *Social risk factor may be statistically significant but does not improve model performance (e.g., C statistic is not improved)*
  - *Effect of social risk factor may often be small*
  - *Access to data can be limited*
- Ongoing concerns about potential differences in quality and the impact on disparities; however, growing evidence in the literature about the impact on access if measures are not adjusted

Standing Committee Discussions

- Continued use of race as a potential variable
  - *Questions about influence of genetics (e.g., varying rate of medication uptake) vs social factors*
  - *Committees indicated a preference for stratification*
- Concerns that social risk factors may be held to a different standard for inclusion
  - *Social risk factor may be statistically significant but does not improve model performance (e.g., C statistic is not improved)*
  - *Concerns that social risk factors are being tested for impact after clinical factors*
- Growing evidence in the literature about the impact on access to care if measures are not adjusted
- Access to data on social risk continues to be a challenge for developers

Methodologies Used for Adjustment and Stratification

- Statistical models and stratification were the most common techniques used in measures submitted for endorsement.
- Developers who used statistical models used various forms of regression analysis:
  - *Hierarchical logistic regression*
  - *Poisson regression*
  - *Ordinary least squares regression (generally the same of linear regression)*
  - *Negative binomial regression*

Methodologies Used for Adjustment and Stratification Continued

There was greater variation in how developers interpreted results and made decisions about which factors to include:

- Rationales for not including:
  - Lack of available data
  - Unable to differentiate patient level or hospital level effect
  - Concerns about masking disparities
  - Factor was significant but small effect size
  - Factor was significant but clinical variables capture the majority of risk
  - Factor was significant but no improvement to model (e.g., c-statistic is unchanged)
- Rationales for including:
  - Factor was significant
  - Hospital level effects not entirely driving results

Current Guidance for Developers

The NQF Measure Developer Guidebook includes instructions for completing the risk-adjustment portion of the measure submission and includes:

- Examples of social risk factors (patient-level, proxy variables, and patient community characteristics)
- Instructions for noting the conceptual rationale that supports/does not support risk adjustment
- Types of analyses that would be appropriate for determining whether a measure should include risk adjustment for social risk factors
- Instructions for comparing performance scores with and without social risk factors in the model
- Request for updated reliability and validity testing if necessary; details of the final statistical risk model; information required to stratify a version of the measure that is clinically adjusted only and the measure results of the social risk variables.

Discussion

- How do you approach risk adjustment for social factors?
  - *Would you be willing to share your method for developing and interpreting adjustment models?*
  - *How do you approach stratification?*
  - *Are there scenarios when process measures should also be considered for adjustment?*
- Which factors do you consider?
  - *Do you feel there is a need for a standardized list of factors to consider?*
- Which data sources are you using to analyze the impact of social risk factors?
  - *Are there data sources you would advise other developers to explore?*

Discussion Continued

- What do you find most challenging about considering and/or testing adjustment for social risk factors?
- How can NQF best guide developers around social risk adjustment?
  - *Should NQF be more prescriptive in methodology or data used?*
- Do you have any initial reflections on the data gathered to date from the trial?
- Do you adhere to the recommendations of the NQF Risk Adjustment for SES Report?

NATIONAL QUALITY FORUM

Measure Sets and Measurement Systems

Erin O'Rourke, Senior Director, NQF
Nicolette Mehas, Director, NQF

The Issues: Current Challenges in Measurement

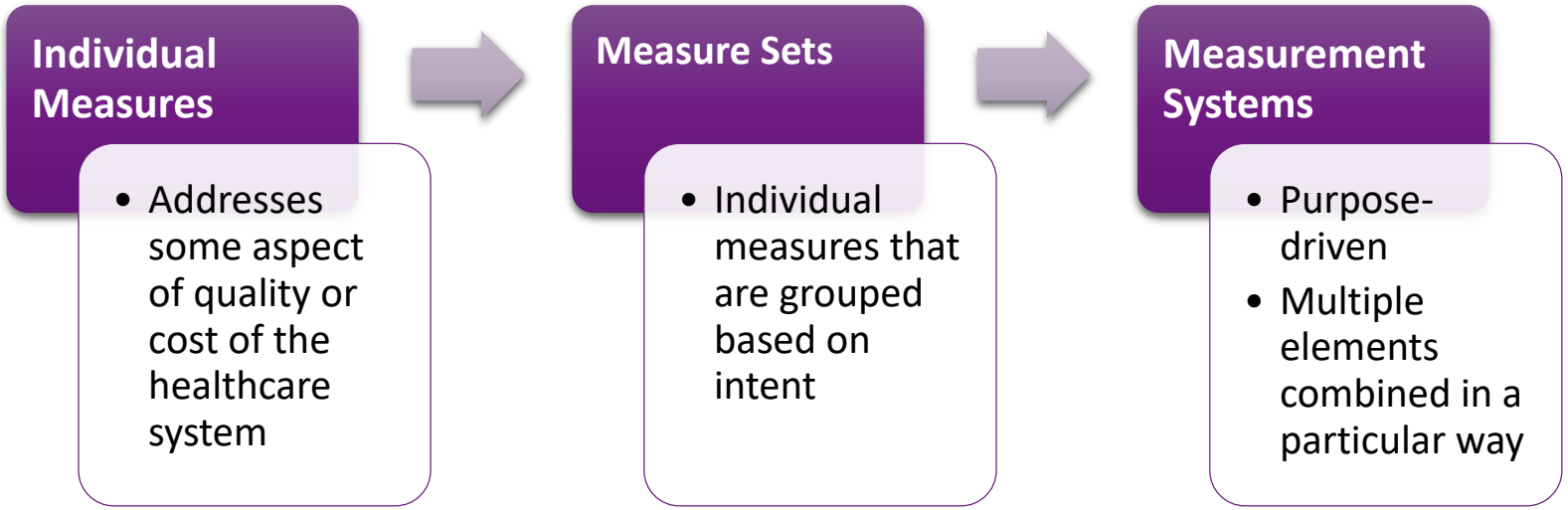
Transition to value

Burden of measurement

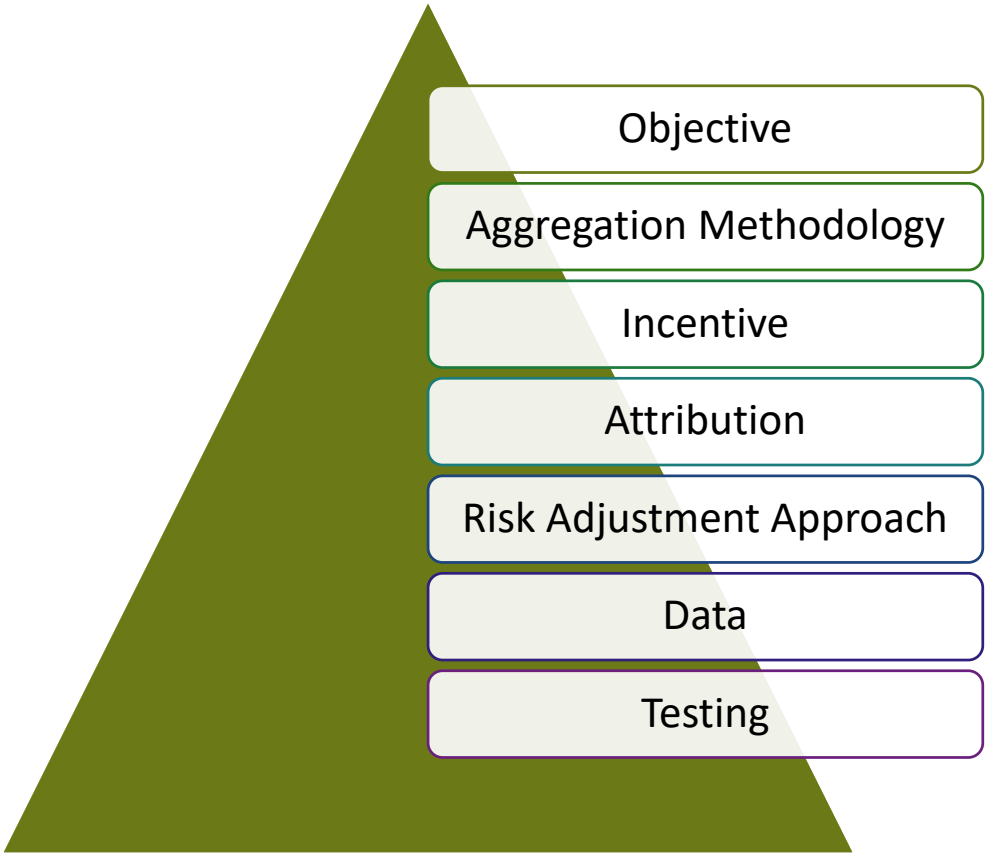
Lack of transparency

The Opportunity: A New Framework for a Quality Measurement Infrastructure

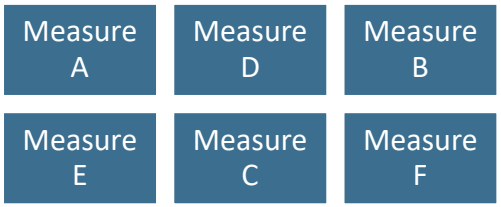
- Establish independent, transparent, and multistakeholder consideration of three interdependent levels of performance measurement:



What is a Measurement System?

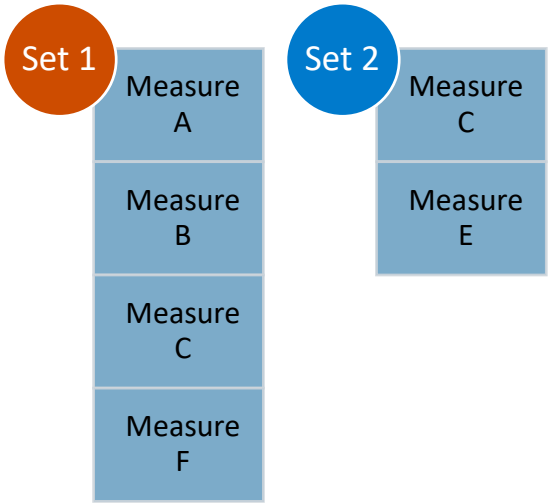


Conceptual Framework: Quality Measurement Infrastructure



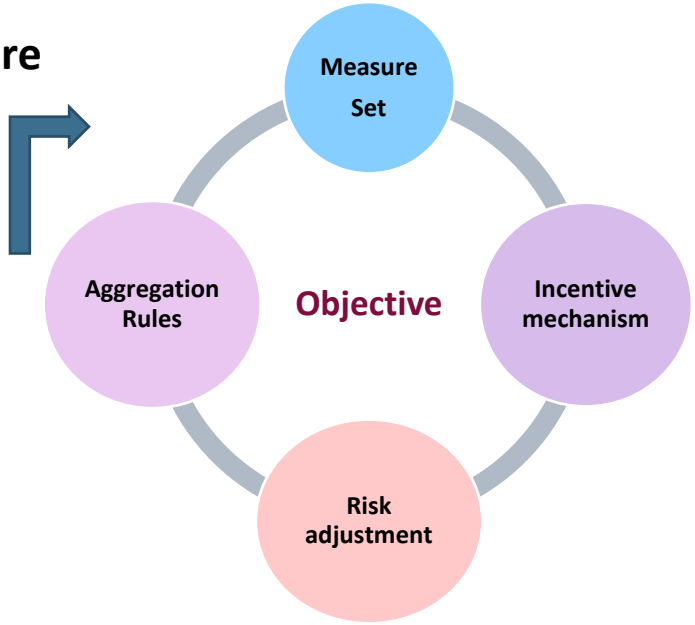
Level 1: Individual Measures

Example: NQF Measure 2158 Medicare Spending Per Beneficiary



Level 2: Measure Sets

Example: CMS Hospital Value-Based Purchasing Program (HVBP) uses a set of measures assessing four domains: safety, clinical care, efficiency and cost reduction, and patient experience.



Level 3: Measurement Systems

Example: Measures are aggregated as part of the HVBP program, and the score is used to determine hospital penalties or rewards.

NQF Work on Measure Sets and Measurement Systems

Measure Sets

- Core Quality Measure Collaborative
- IHA Benchmarking
- Shatterproof Rating System for Addiction Treatment Programs Measure Set
- MAP Rural Health Measure Set

Measurement Systems

- Harvard Medical School/Laura and John Arnold Foundation project
- NQF Measurement Systems

Conclusion of Phase 1—Measurement Systems

- NQF/Harvard Medical School white paper published on March 20
- NQF Annual Conference talks and panels
  - *Pioneering Pathways to Value*
    - Elisa Munthali, MPH, Senior Vice President, Quality Measurement, National Quality Forum
  - *New Approaches to Measurement – Measurement Systems*
    - Michael Chernew, PhD, Leonard D. Schaeffer Professor of Health Care Policy and the director of the Healthcare Markets and Regulation (HMR) Lab in the Department of Health Care Policy at Harvard Medical School
    - Jeffrey Rideout, MD, President and CEO, IHA
    - Gary Mendell, CEO, Shatterproof
    - Moderator - Aisha Pittman, MPH, Senior Director, Premier

Phase 2 — Project Overview

- Present-March 2020
- Convene a TEP (First meeting, late June)
  - *Develop standardized definitions*
  - *Define best practices and principles*
  - *Discuss data issues*
  - *Identify unintended consequences*
- Ten web meetings
- Two Reports
  - *Draft and final report on best practices and principles for measure sets*
  - *Draft and final report on best practices and principles for measurement systems*

Phase 2 – Evaluation of Measure Sets and Measurement Systems

- NQF Annual Conference
  - Present measure sets and measurement systems at the Annual Conference
- TEP Call for Nominations
- Joint Orientation
- Begin Measure Sets work
- Begin Measurement Systems work
- Publish draft recommendation for Measure Sets
- Public comment on draft recommendations for Measure Sets
- Finalize recommendations for Measure Sets
- Continue Measurement Systems work
- Publish draft recommendations for Measurement Systems
- Public commenting period on draft recommendations for Measurement Systems
- Finalize recommendations for Measurement Systems

Draft Definitions

Measure Set – a collection of individual performance measures that address an aspect of cost or quality, often grouped based on intent

Components to consider:

- Intent/grouping
- Selection criteria and process
- Application (e.g., need to implement entire set versus ability to select certain measure)
- Maintenance
- Feedback/evaluation

Draft Definitions

Measurement System – a group of individual measures that, based on a predefined methodology, work together to assess quality or cost in relationship to a goal

Components to consider:

- Intent
- Aggregation and weighting methods
- Attribution methodology
- Data
- Risk adjustment
- Evaluation methods

Discussion

- How do you think about measure sets and measurement systems?
- What has been your experience in grouping measures by concept or based on intent?
- Is there a potential impact on the measure development process if measures are designed to be used in a set or system versus used individually?
- Are there considerations from the measure developer perspective that the TEP should consider?
- What are your overall thoughts on how sets and systems should be designed and evaluated?

Break

NATIONAL QUALITY FORUM

Human-Centered Design in Practice for Measurement Development

Jayanti (Jay) Bandyopadhyay, Mathematica Policy Research

Human-Centered Design in Practice for Measure Development

Generating measure concepts for pain management using the patient experience

Presentation at the National Quality Forum
Measure Developer Conference

Jayanti Bandyopadhyay

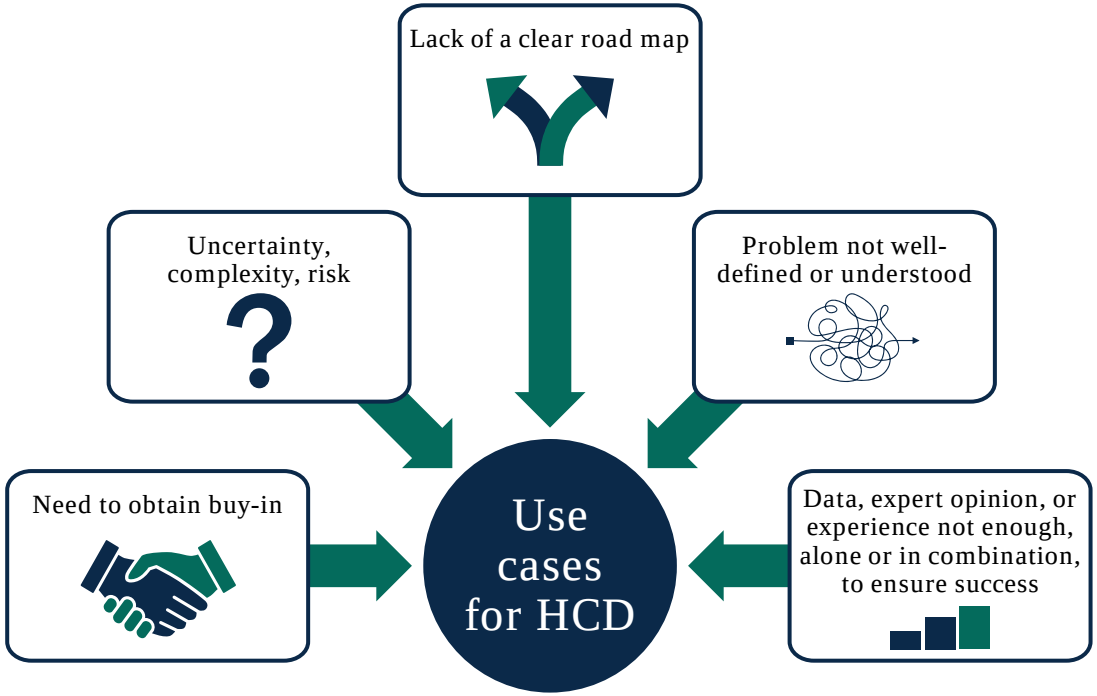
June 10, 2019

Agenda

- **What is human-centered design (HCD)?**
- **Why is it important?**
- **How did we use it?**
- **What happened during the phases of work?**
- **What did we learn and why is it important to measure development?**

What Is HCD?

When Can We Use HCD?

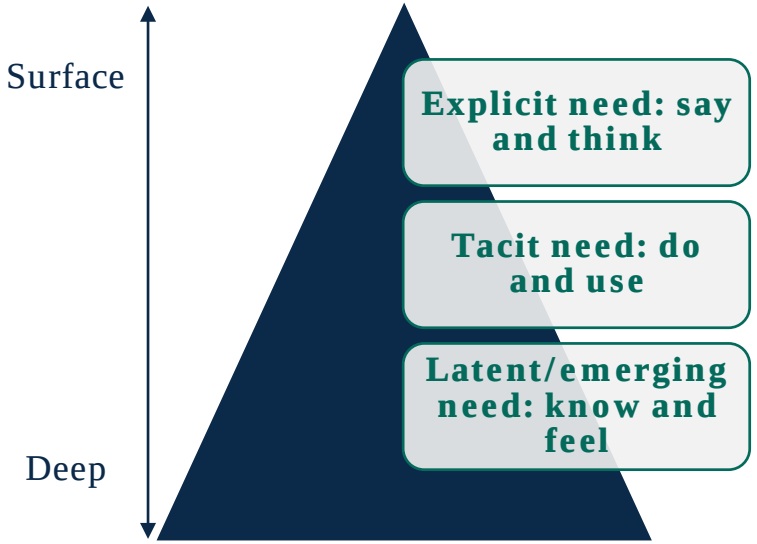


Why Is HCD Important?

- **HCD is a highly collaborative method to collect deep user insights**
- **CMS is emphasizing increased use of HCD as it is a "sibling" of patient-centered care**

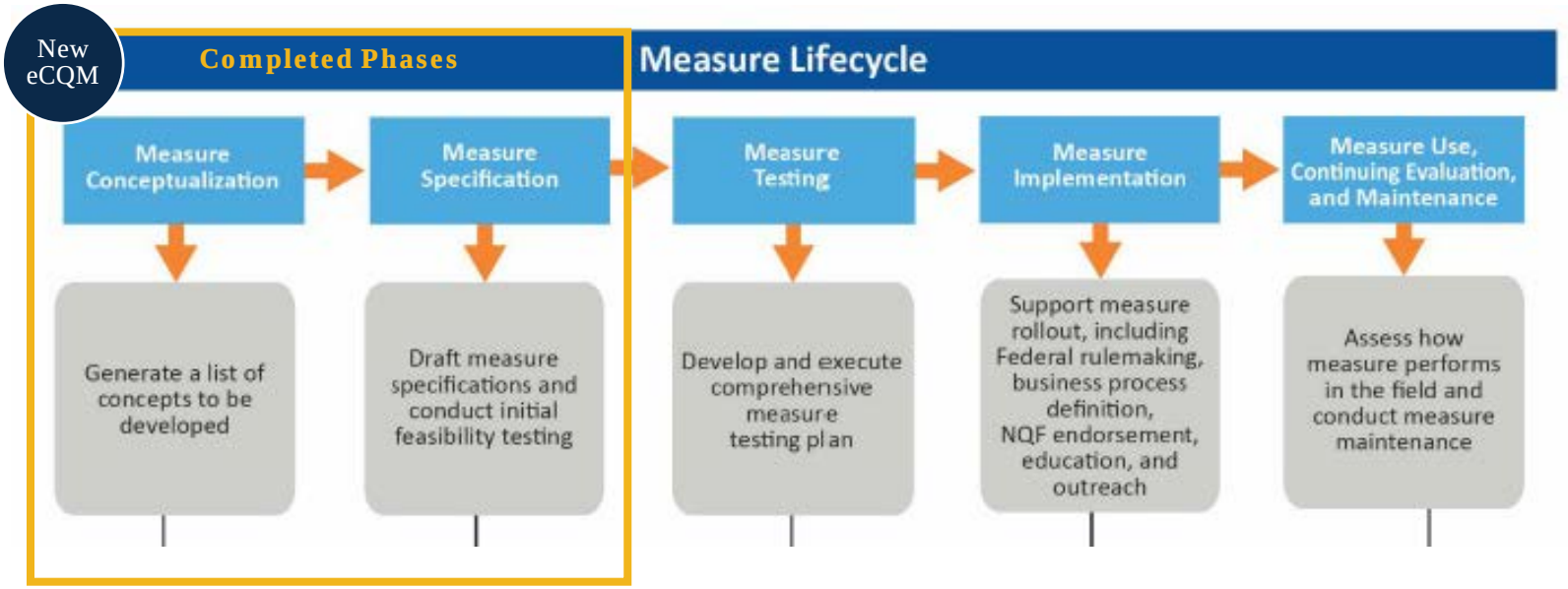
"I'd assess things myself and say, it's been a couple of days, how is it feeling, is it getting any better or is it worse? Does this seem to be a longer-term thing? Is ice helping? Is heat helping? Are anti-inflammatories helping? Is this a sharp pain? Is it a dull pain? How often is this occurring? How much function have I lost? Are things getting better over a little time?"

—Interview participant with chronic pain

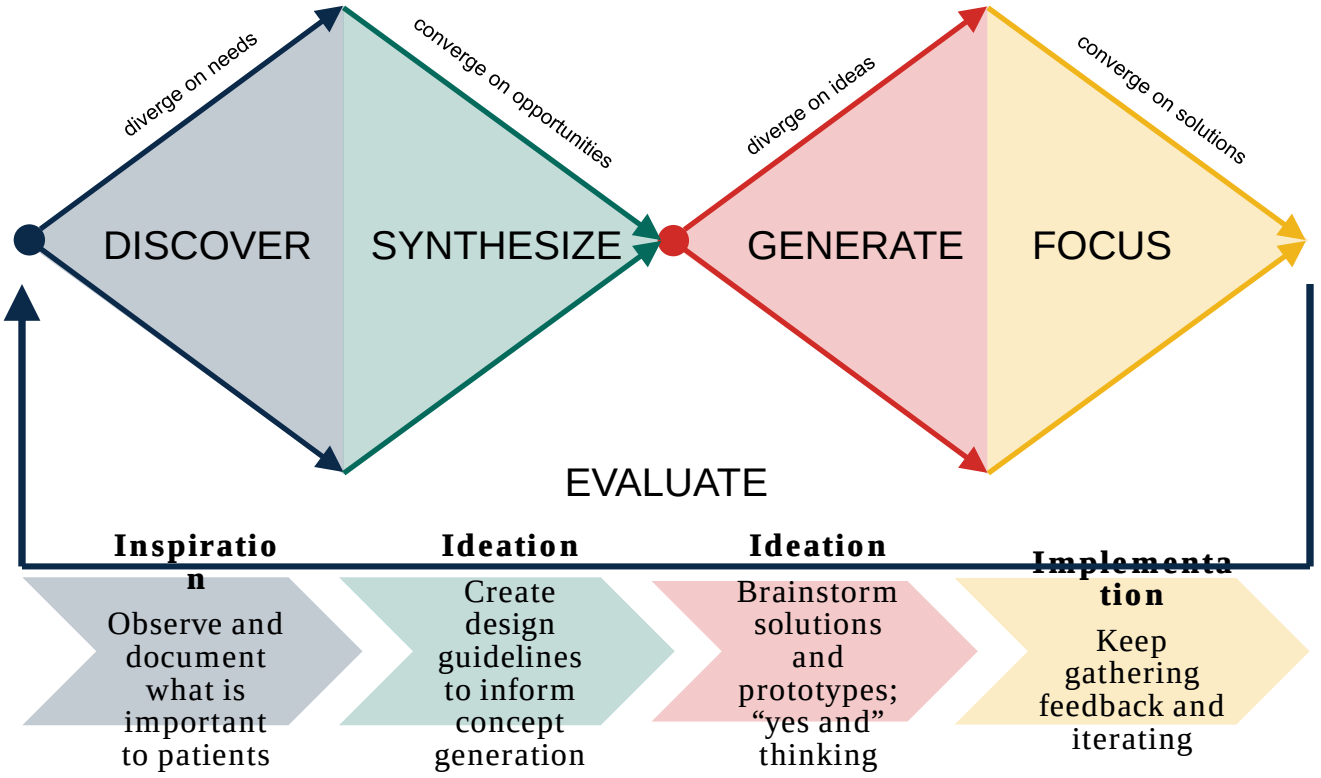


How Did We Use HCD in Our Work?

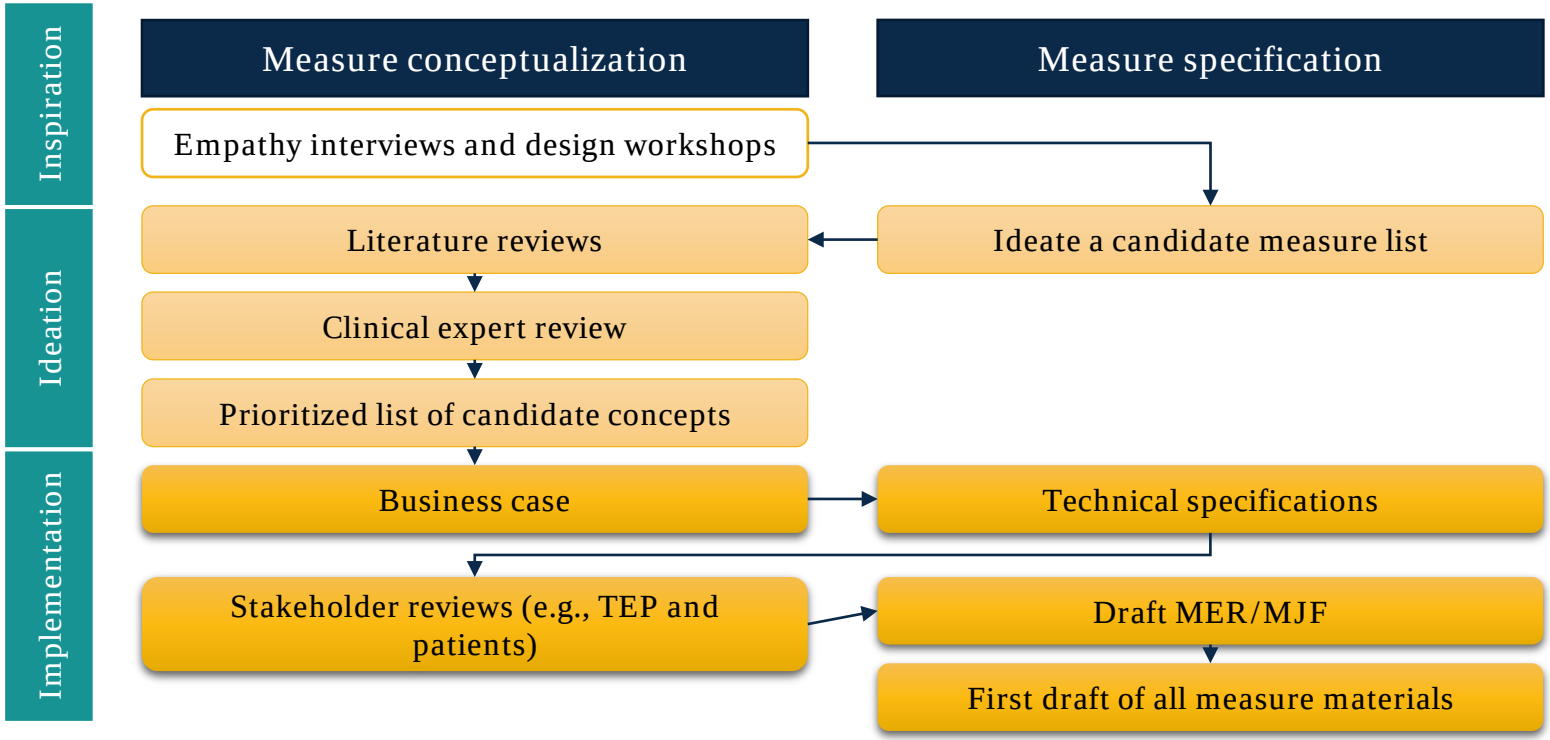
CMS MMS Blueprint Measure Lifecycle



HCD Phases


diverge on needs
converge on opportunities
diverge on ideas
converge on solutions
DISCOVER
SYNTHESIZE
GENERATE
FOCUS
EVALUATE
Inspiration
Observe and document what is important to patients
Ideation
Create design guidelines to inform concept generation
Ideation
Brainstorm solutions and prototypes; "yes and" thinking
Implementation
Keep gathering feedback and iterating

Integrating HCD into the Measure Lifecycle

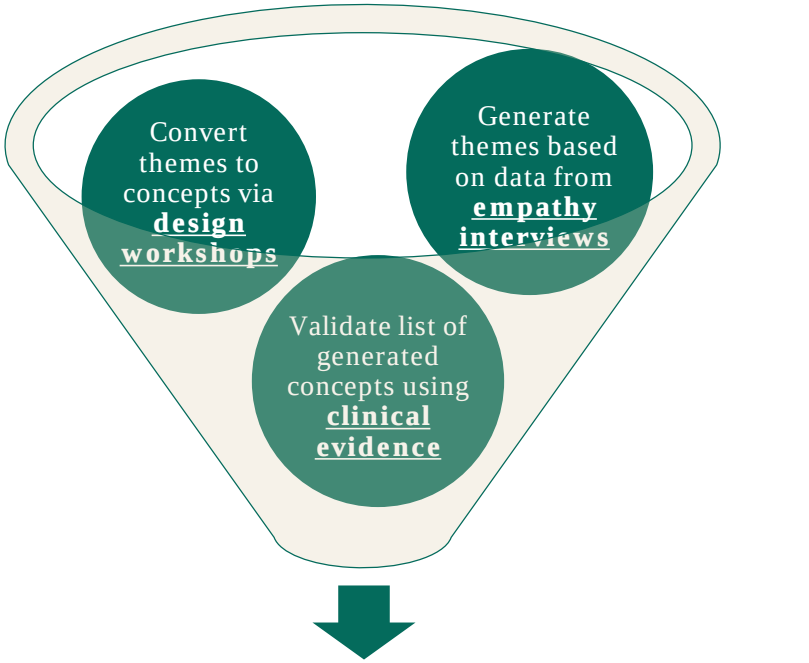


Case Study: Developing a Measure Concept for Pain Management

Context

- **Objective: Develop a measure to balance potential unintended consequences of Potential Opioid Overuse measure that identifies unmet needs of patients experiencing pain**
- **Approach: Use HCD to explore patients' needs**
- **Key considerations: Is there a gap between patient priorities and existing clinical evidence?**

Revised Information-Gathering Approach



Case Study Road Map

- Inspiration: Gather information on what is important to patients
- Ideation: Create design guidelines that will inform generation of measure concepts
- Ideation: Validate themes from empathy interviews and ideate solutions and concepts with patients and clinicians

What Were the Inputs and Outputs of Each Phase?

Phase 1: Empathy Interviews

What: Because pain is a complex biopsychosocial issue, we need to better understand the target population (patients with acute or chronic pain) and empathize with their experience, including their mindsets, behaviors, lifestyles, desires, fears, and opinions

- Measure lifecycle phase: conceptualization
- Human-centered design phase: inspiration

Phase 1: Empathy Interviews—Input

Why (informed by research questions):

- Understand the health outcomes most important to patients
- Identify patients' unmet or unarticulated needs regarding opioids
- Explore how patients define "quality of care"
- Assess the face validity of consensus guideline recommendations against the patient perspective

How:

- Conducted one-hour, semistructured, one-on-one interviews with patient participants

Phase 1: Empathy Interviews—Output

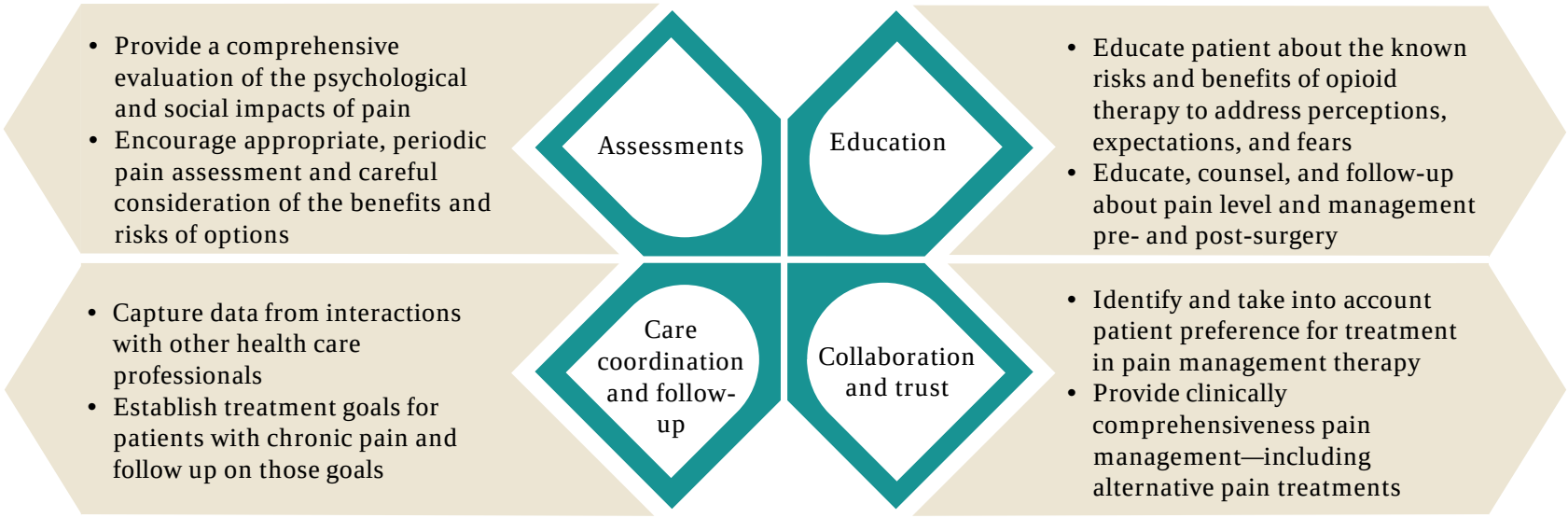
"I think, because of the pain, my muscles tightened up. . . . So that's where the massage came in, and I tried to swim. Just certain things for the muscle part, because I learned that they were interconnected. . . . So I think those alternative therapies, I guess, sort of offset the way my body reacted to the pain."

—Interview participant with chronic pain

Participants used alternative therapies as their primary source of pain treatment, resorting to opioids only when necessary

The concept/measure should consider whether the provider has already recommended alternative therapies and/or the patient has already tried them

Phase 1: Empathy Interviews—Output



Phase 2: Design Workshops—Input

What: Create a draft list of concepts via design workshops

- Measure lifecycle phase: specification
- Human-centered design phase: ideation

Why (informed by research questions):

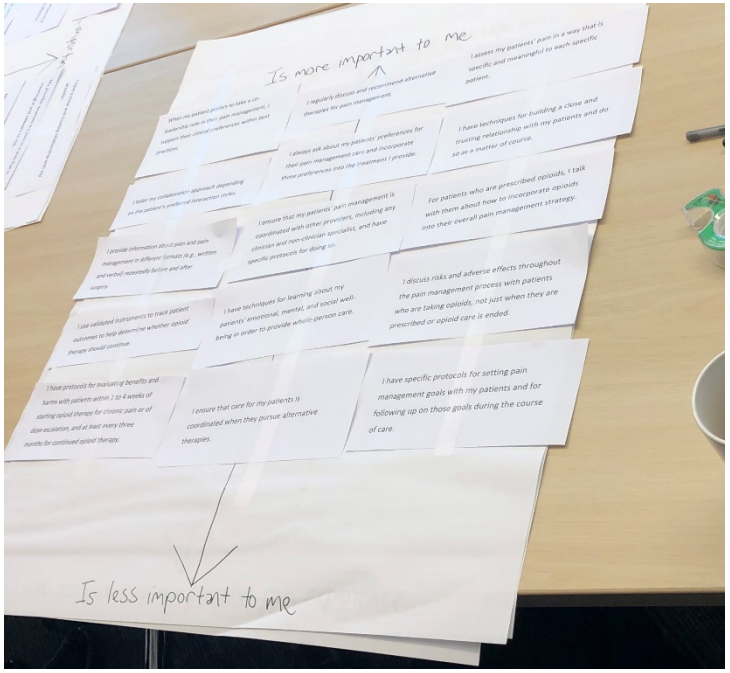
- Assess validity of the themes from empathy interviews
- Draft list of initial measure concepts

How:

- Conducted two three-hour design workshops with eight new patient participants and two clinician participants
- During data analysis, fleshed out the list of measure concepts drafted and prioritized by patients and clinicians

Phase 2: Output—Prioritization

Conducting a prioritization activity with the patient and clinician participants allowed us to understand the most important aspects of the pain management from the themes and insights we gained from the empathy interviews.



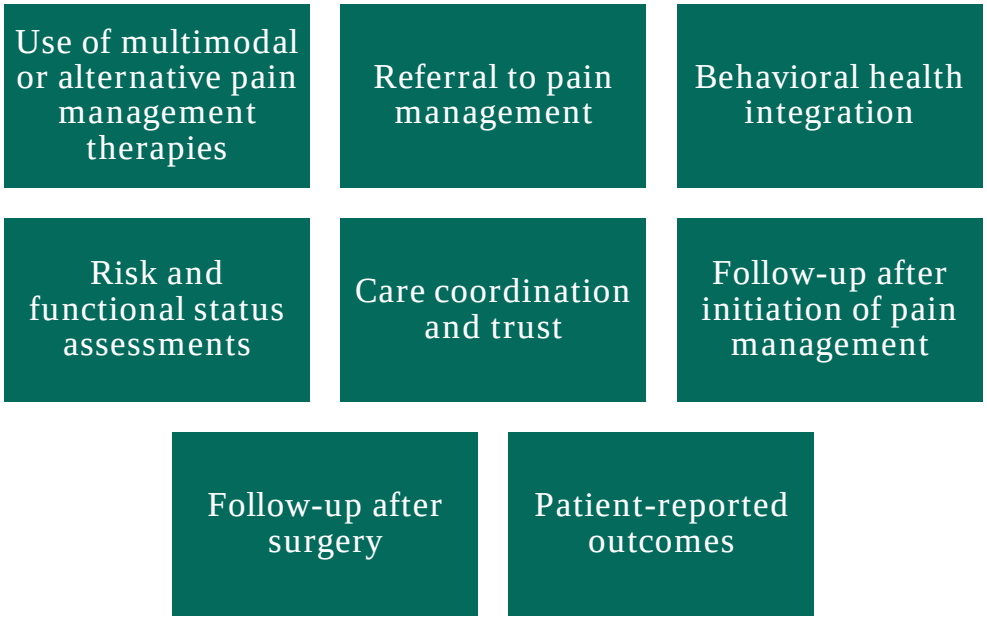
Phase 2: Output—Vision Board Collage

Another activity we used with the patient and clinician participants was the vision board collage, where we asked them to collage their ideal pain management experience. Using this method allows participants to suggest possible solutions, engaging them directly in problem-solving, and is particularly helpful during the concept ideation phase.


PRIORITIZATION
COLLAGING
Love Letters/Breakup Letters

Phase 2: Output—Initial List of Measure Concepts

We ideated 44 new potential measure concepts, categorized into 8 high-level domains based on the underlying topics (intervention, process, or outcome) of the concepts.



Phase 3: Literature Review and Clinician Input

What: Validate list of generated concepts using evidence from literature review and stakeholder input

- Measure lifecycle phase: conceptualization

Why:

- Determine which concepts have a strong evidence base
- Gain input on the face validity, feasibility (Expert Workgroup only), and usability of the concepts to prioritize them

How:

- Convene opioid EWG
- Conduct literature review

Phase 3: Output—Top Concept

- **Domain: Use of multimodal or alternative pain management therapies**
- **Patient input: Most patients we interviewed identified pain management therapies other than opioid therapy to be their preference for first-line therapy**
- **Expert input: Clinicians who participated in the design workshops noted that they also drew on nonopioid approaches to pain management**

What Did We Learn by Using HCD in Practice for Measure Development?

Benefits and Outcome of Innovative HCD Approach

Why now?

- **HCD is an excellent fit for data collection in the measure conceptualization phase**

- **We were able to identify hidden insights or unknowns to patients and measure developers**

- **HCD is also an excellent fit for analyzing all the qualitative data in a very short turnaround**

- **Potential unintended consequences or implementation challenges are brought to the forefront**

Benefits and Outcome of Innovative HCD Approach

Findings specific to our case:

- **Most concepts identified as important from patient perspective aligned with clinician consensus statements and national guidelines**
  - For example, patients did not like to use opioids as first-line therapy, which the CDC guidelines also recommend

- **Patients and clinicians felt very engaged; brought them together to ideate on concepts**

What's Next?

- **User testing via patient reengagement**
- **Field testing the measure**
- **Continued development using testing and stakeholder feedback**

For more information

Jayanti Bandyopadhyay

JBandyopadhyay@mathematica-mpr.com

Cindy Cullen

CCullen@mathematica-mpr.com

Sam Simon

SSimon@mathematica-mpr.com

NATIONAL QUALITY FORUM

Measure Feedback Loop

Jean-Luc Tilly, Senior Data Analytics Manager, NQF

Welcome and Introductions

NQF Project Staff

- Ashlie Wilbon, MS, MPH, FNP-C, Senior Director
- Jean-Luc Tilly, Senior Manager, Data Analytics
- Madison Jung, Project Manager
- Navya Kumar, MPH, Project Analyst

Session Objectives

- Review Findings and Recommendations of Feedback Loop Committee
- Identify new opportunities for collaboration on generating and analyzing feedback on performance measures

Measure Feedback Loop Committee

- **Co-chair:** Rose Baez, RN, MSN, CPHQ, CPPS
- **Co-chair:** Edison Machado, MD, MBA
- Constance Anderson, BSN, MBA
- Robert Centor, MD, MACP
- Elvia Chavarria, MPH
- Dan Culica, MD, PhD
- Melody Danko Holsomback
- Anne Deutsch, RN, PhD
- Tricia Elliott, MBA, CPHQ
- Lee Fleisher, MD
- Mark E. Huang, MD
- Joseph Kunisch, PhD, RN-BC, CPHQ
- Claire Noel-Miller, MPA, PhD
- Ekta Punwani, MHA
- Koryn Rubin, MHA
- Elizabeth (Beth) Rubinstein
- Sue Sheridan, MIM, MBA, DHL
- Jill Shuemaker, RN, CPHIMS
- Heather Smith, PT, MPH
- Deborah Struth, MSN, RN, PhD(c)
- Sara Toomey, MD, MPhil, MPH, MSc

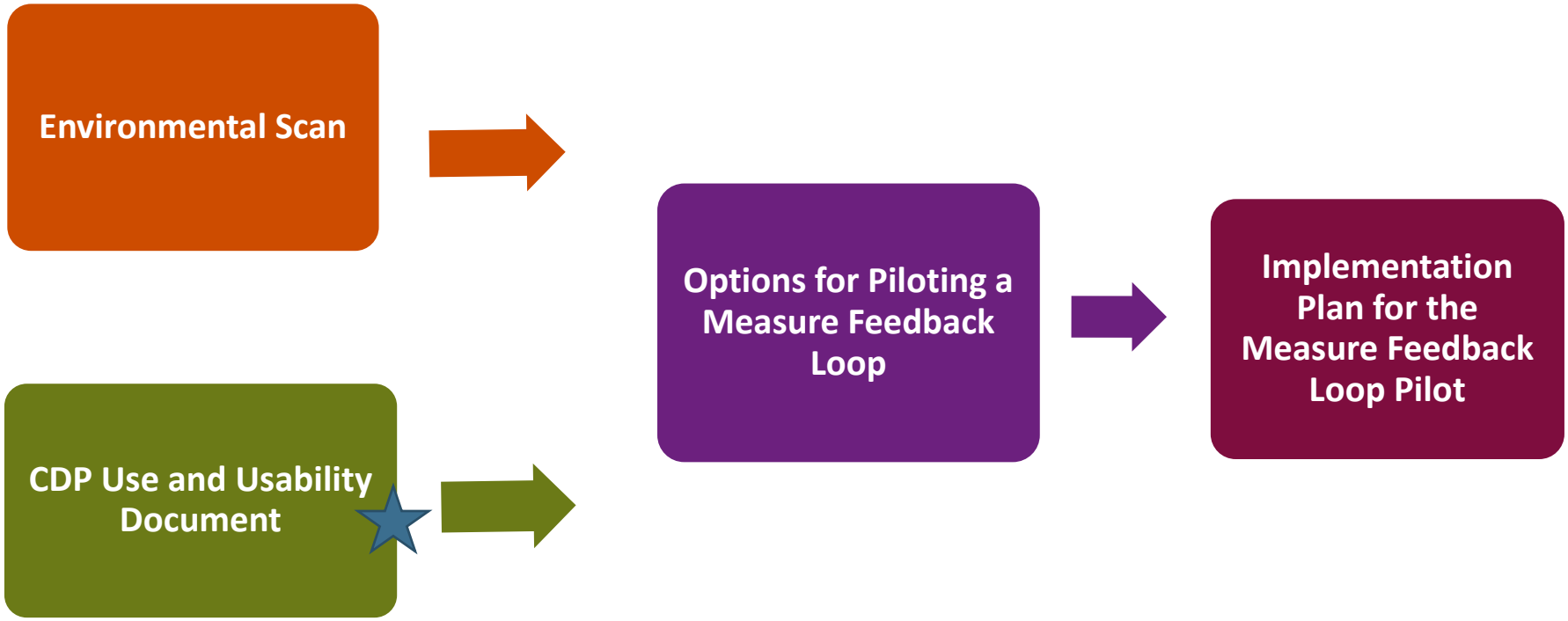
Federal Liaisons

CMS

- Maria Durham
- Sophia Chan
- Patrick Wynne
- Melissa Evans

Project Overview

Project Deliverables



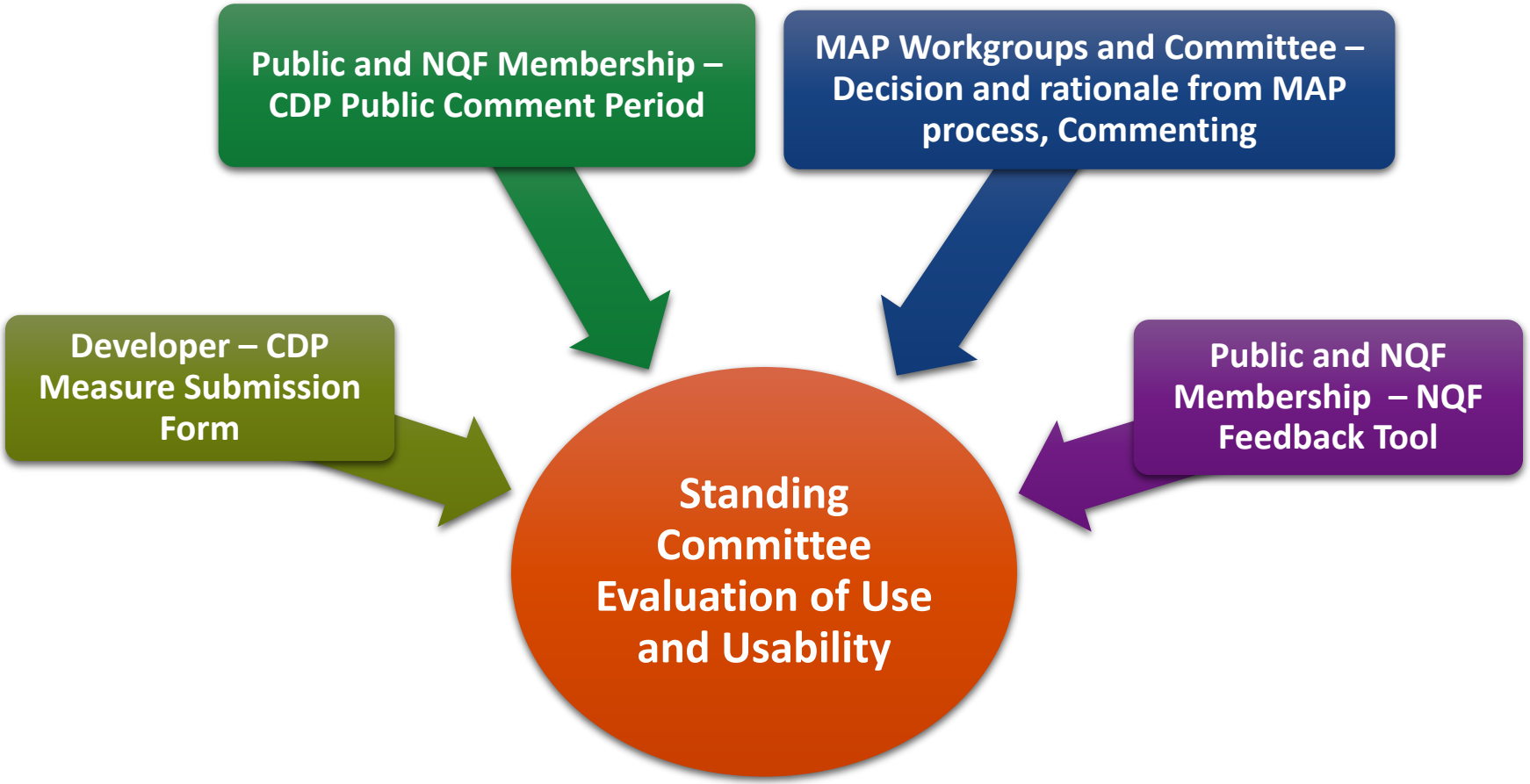
Project Overview

- Objectives
  - To understand outcomes, and what the unintended consequences are, if any
  - To understand how a measure actually performs when in use, and what the possible issues or risks are that may be associated with measure implementation
  - To help address whether the measure is having its intended effects on improving quality of care and health measurement
- Definitions
  - Feedback loop
    - » Refers to the process by which feedback from the measure is relayed back to the multistakeholder standing committee members who recommended the measure to be (re-) endorsed or selected for program use.
    - » In previous CDP projects, standing committee members have expressed the need for updates on how a measure has performed after endorsement. This is especially the case for measures that are contentious, and have a chance of impacting certain stakeholders negatively.
  - Feedback
    - » Refers to information about measure performance that could be based on quantitative data or qualitative information

Evaluation of Measures Requires Feedback

- Importance
  - *Assessment of performance gap, opportunity for improvement*
- Feasibility
  - *Assessment of whether there are any significant barriers to implementation*
- Usability and Use
  - *Assessment of whether the measure is (or will be) in use in an accountability application*
  - *Assessment of unintended consequences and benefits of the measure*
  - *Assessment of feedback received on the measure and how it was used*

Feedback is Considered from Multiple Perspectives and Channels during Evaluation



Environmental Scan Report — Challenges

- Variation in Data Collection Processes
  - *Lack of formal to collect feedback across stakeholders*
  - *Different stakeholders and types of measures generate different feedback, making consolidating difficult*
- Timing of Feedback
  - *Feedback received too late in development cycle*
  - *Misalignment of deadlines in the measure lifecycle*
- Burden
  - *Difficult to determine if a measure is meaningful and without unintended consequences*
  - *Lack of knowledge about to find and send measure feedback in one central source*

Use and Usability Report—Recommendations

- Create an Opportunity for Exceptions to “Use” criterion
- Improve Access and Opportunities to Submit Feedback
- Close the Feedback Loop
- Target Outreach to Key Stakeholders
- Classify Feedback
- Develop Guidance for Developers in Feedback Collection

Discussion Questions

- What role should developers play in the feedback loop?
- What strategies does your organization use to collect feedback on your measures? How often is this feedback collected? From whom? What channels (e.g., commenting, office hours)?
- What strategies does your organization use to determine who is using your measure?
- What are reasonable parameters for response time to feedback?
- How can developers and NQF partner to identify new sources of measure feedback? Other sources for soliciting feedback?
- What kinds of feedback warrant a change, or a maintenance of endorsement review?
- Are you familiar with the NQF measure feedback tool? What would make this tool more useful to your organization in collecting feedback on your endorsed measures?

Feedback Tool

- [screenshare]

Overview of Meeting Timeline

Meeting	Date
Web Meeting 3 and 4: Measure Feedback and the NQF CDP Process, Part 1 and 2 [3 hours each]	April 30, 2019, 2-5 pm ET May 7, 2019, 2-5 pm ET
Web Meeting 5: Options for Piloting the Measure Feedback Loop, Part 1 [2 hours]	July 24, 2019, 1-3 pm ET
Web Meeting 6 and 7: Options for Piloting the Measure Feedback Loop, Parts 2 and 3 [2 hours each]	September 3, 2019, 2-4 pm ET September 5, 2019, 2-4 pm ET
Web Meeting 8: Implementation Plan [2 hours]	November 19, 2019, 2-4 pm ET
Web Meeting 9: Project Wrap-Up [2 hours]	January 16, 2020, 1-3 pm ET

Project Contact Information

- Email: measurefeedback@qualityforum.org
- NQF phone: 202-783-1300
- Project page: https://www.qualityforum.org/Measure_Feedback_Loop.aspx
- SharePoint: http://share.qualityforum.org/Projects/MeasureFeedbackLoop/SitePages/Home.aspx

NATIONAL QUALITY FORUM

eCQM Submission Considerations

Chris Millet, NQF Consultant

1. Data Elements

- Requires all data elements used in eCQM
- Not just “critical” data elements*

*supplemental data elements

What do we mean by “all” data elements used in the eCQM specification?

Data Criteria (QDM Data Elements)

- "Diagnosis: Bipolar Diagnosis" using "Bipolar Diagnosis (2.16.840.1.113883.3.600.450)"
- "Diagnosis: Depression diagnosis" using "Depression diagnosis (2.16.840.1.113883.3.600.145)"
- "Encounter, Performed: Depression Screening Encounter Codes" using "Depression Screening Encounter Codes (2.16.840.1.113883.3.600.1916)"
- "Intervention, Order: Referral for Depression Adolescent" using "Referral for Depression Adolescent (2.16.840.1.113883.3.600.537)"
- "Intervention, Order: Referral for Depression Adult" using "Referral for Depression Adult (2.16.840.1.113883.3.600.538)"
- "Intervention, Performed: Additional evaluation for depression - adolescent" using "Additional evaluation for depression - adolescent (2.16.840.1.113883.3.600.1542)"
- "Intervention, Performed: Additional evaluation for depression - adult" using "Additional evaluation for depression - adult (2.16.840.1.113883.3.600.1545)"
- "Intervention, Performed: Follow-up for depression - adolescent" using "Follow-up for depression - adolescent (2.16.840.1.113883.3.600.467)"
- "Intervention, Performed: Follow-up for depression - adult" using "Follow-up for depression - adult (2.16.840.1.113883.3.600.468)"
- "Intervention, Performed: Suicide Risk Assessment" using "Suicide Risk Assessment (2.16.840.1.113883.3.600.559)"
- "Medication, Order: Depression medications - adolescent" using "Depression medications - adolescent (2.16.840.1.113883.3.600.469)"
- "Medication, Order: Depression medications - adult" using "Depression medications - adult (2.16.840.1.113883.3.600.470)"
- "Patient Characteristic Ethnicity: Ethnicity" using "Ethnicity (2.16.840.1.114222.4.11.837)"
- "Patient Characteristic Payer: Payer" using "Payer (2.16.840.1.114222.4.11.3591)"
- "Patient Characteristic Race: Race" using "Race (2.16.840.1.114222.4.11.836)"
- "Patient Characteristic Sex: ONC Administrative Sex" using "ONC Administrative Sex (2.16.840.1.113762.1.4.1)"
- "Assessment, Not Performed: Adolescent depression screening assessment" using "Adolescent depression screening assessment (LOINC version 2.63 Code 73831-0)"
- "Assessment, Not Performed: Adult depression screening assessment" using "Adult depression screening assessment (LOINC version 2.63 Code 73832-8)"
- "Assessment, Performed: Adolescent depression screening assessment" using "Adolescent depression screening assessment (LOINC version 2.63 Code 73831-0)"
- "Assessment, Performed: Adult depression screening assessment" using "Adult depression screening assessment (LOINC version 2.63 Code 73832-8)"

All data elements should be listed in the Feasibility Scorecard

LIST ALL DATA ELEMENTS - this will pre-populate scorecards	
	Data Element
1	"Diagnosis: Bipolar Diagnosis" using "Bipolar Diagnosis (2.16.840.1.113883.3.600.450)"
2	"Diagnosis: Depression diagnosis" using "Depression diagnosis (2.16.840.1.113883.3.600.145)"
3	"Encounter, Performed: Depression Screening Encounter Codes" using "Depression Screening Encounter Codes
4	"Intervention, Order: Referral for Depression Adolescent" using "Referral for Depression Adolescent (2.16.840.
5	"Intervention, Order: Referral for Depression Adult" using "Referral for Depression Adult (2.16.840.1.113883.3
6	"Intervention, Performed: Additional evaluation for depression - adolescent" using "Additional evaluation for de
7	"Intervention, Performed: Additional evaluation for depression - adult" using "Additional evaluation for depress
8	"Intervention, Performed: Follow-up for depression - adult" using "Follow-up for depression - adult (2.16.840.1
9	"Intervention, Performed: Suicide Risk Assessment" using "Suicide Risk Assessment (2.16.840.1.113883.3.600
10	"Medication, Order: Depression medications - adolescent" using "Depression medications - adolescent (2.16.84
11	"Medication, Order: Depression medications - adult" using "Depression medications - adult (2.16.840.1.11388
12	"Patient Characteristic Ethnicity: Ethnicity" using "Ethnicity (2.16.840.1.114222.4.11.837)"
13	"Patient Characteristic Payer: Payer" using "Payer (2.16.840.1.114222.4.11.3591)"
14	"Patient Characteristic Race: Race" using "Race (2.16.840.1.114222.4.11.836)"
15	"Patient Characteristic Sex: ONC Administrative Sex" using "ONC Administrative Sex (2.16.840.1.113762.1.4.1
16	"Assessment, Not Performed: Adolescent depression screening assessment" using "Adolescent depression scre
17	"Assessment, Not Performed: Adult depression screening assessment" using "Adult depression screening asses
18	"Assessment, Performed: Adolescent depression screening assessment" using "Adolescent depression screenin
19	"Assessment, Performed: Adult depression screening assessment" using "Adult depression screening assessme
20	

Exception: Data element attributes

LIST ALL DATA ELEMENTS
Encounter, performed: SampleEncounterValueSet
Encounter, performed: SampleEncounterValueSet admission source
Encounter, performed: SampleEncounterValueSet date/time

How do we review missing data elements?

- Staff analysis to the committee includes a list of all low scoring data elements and will:
  - Highlight which domains were low scoring
  - Include any developer comments or justification

- For missing data elements, staff will:
  - List of all data elements that were not assessed
  - **Recommend a insufficient rating on Feasibility**

Tip

IF a data element is listed in the eCQM Data Criteria section,

THEN It should be listed in Feasibility Scorecard

2. Summary/Average Scores

SUMMARY CURRENT	Data availability	Data accuracy	Data standards	Workflow
Sum of Scores	35	35	36	36
Average within Domain	2.9	2.9	3.0	3.0
Data Elements Scoring Three with Domain	11	11	12	12
Total data elements	12	12	12	12
Percent of data elements currently feasible within domain	92%	92%	100%	100%

2. Summary/Average Scores

SUMMARY CURRENT
Sum of Scores
Average within Domain
Data Elements Scoring Three with Domain
Total data elements
Percent of data elements currently feasible within domain

Any Feasibility Issues

- **ANY** data element scored less than 3 at **ANY** site

3. Current vs Future

Data Element	Timeline	DATA AVAILABILITY: Is the data readily avaiable in a structured format?	
		Score	Additional Characteristics
Encounter, performed: SampleEncounterValueSet	Current	2	
	Future	3	
Encounter characteristic: admission date and time	Current	3	

3. Current vs Future

	DATA AVAILABILITY	
Timeline	Is the data readily avaiable in a structured format?	
	Score	Additional Characteristics
Current	2	
Future	3	

Summary

- **Assess every data element**
- **For any feasibility issues**
- **If implemented right now**

Summary of the Day and Next Steps

Adjourn