NATIONAL QUALITY FORUM

Measure Developer Workshop

National Quality Forum

May 17, 2017

NATIONAL QUALITY FORUM

Measure Prioritization & Graduated Measurement

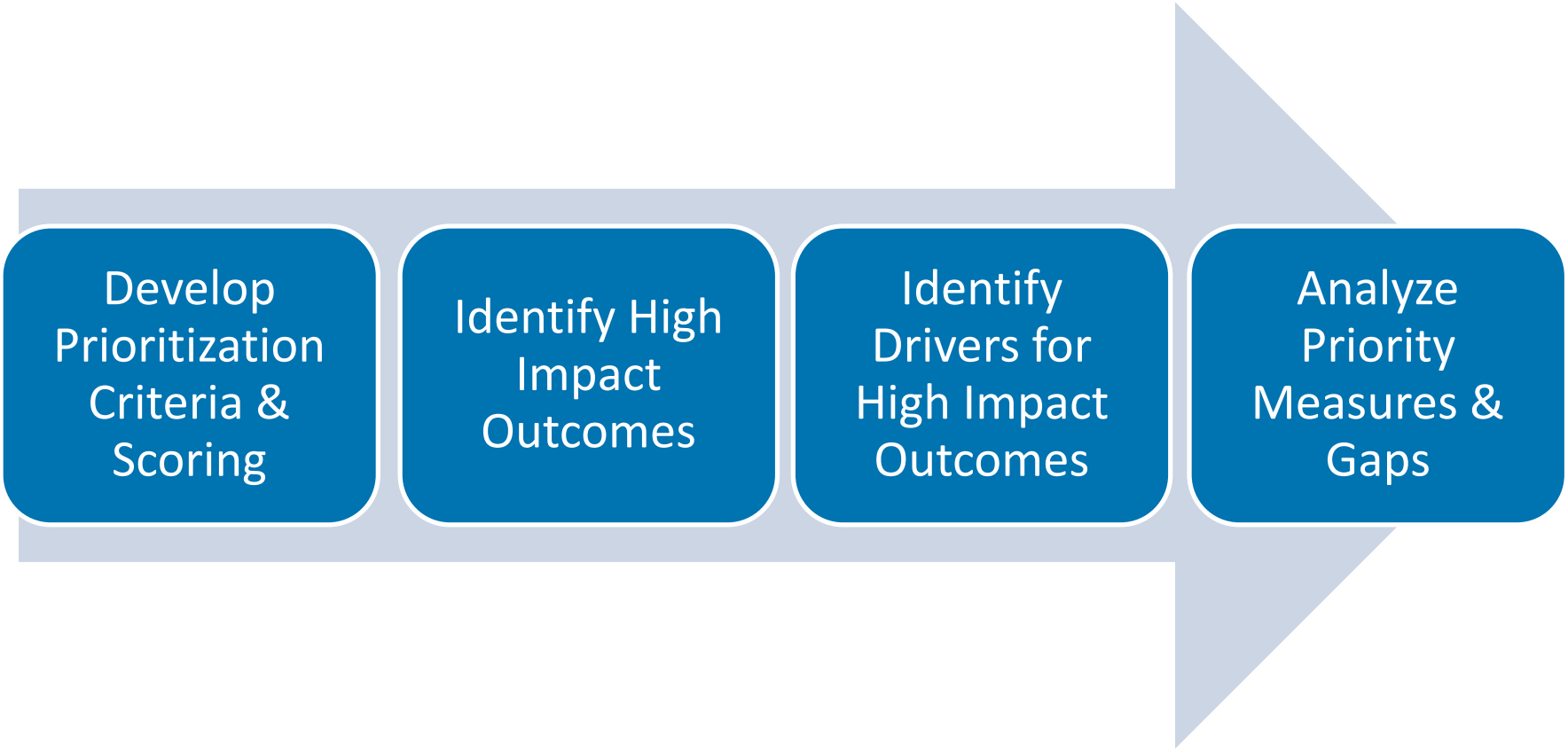
Dr. Helen Burstin, Chief Scientific Officer, NQF
Elisa Munthali, Vice President – Quality Measurement, NQF

May 17, 2017

NQF Strategic Vision

Prioritize Measures

NQF Measure Prioritization Process



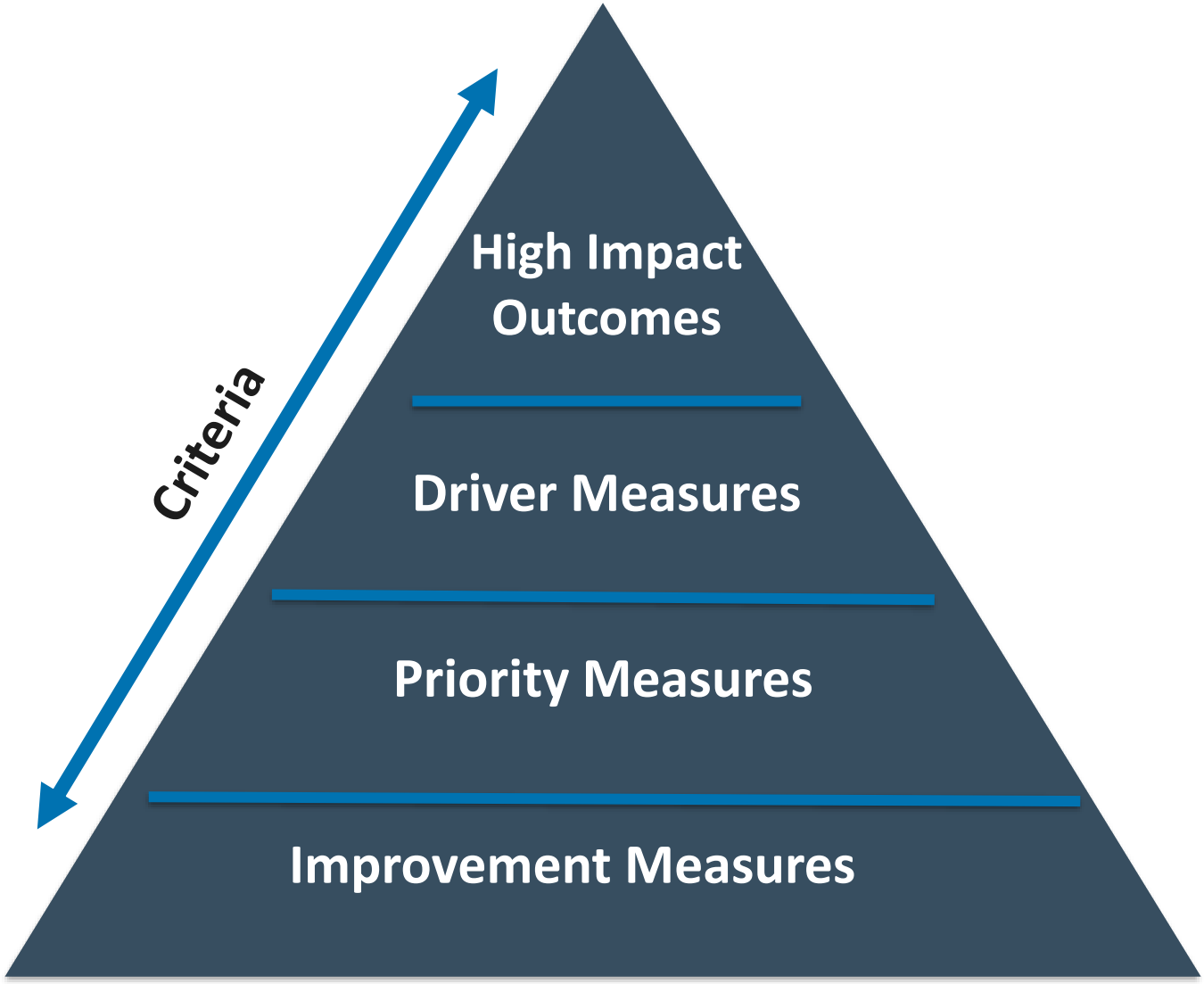
Prioritization Criteria: Environmental Scan

- National Quality Strategy
- IOM Vital Signs
- NQF Prioritization Advisory Committees
- Healthy People 2020 Indicators
- Kaiser Family Foundation Health Tracker
- Consumer priorities for Hospital QI and Implications for Public Reporting, 2011
- IOM: Future Directions for National Healthcare Quality and Disparities Report, 2010
- IHI Whole System Measures
- Commonwealth Fund International Profiles of Healthcare Systems, 2015
- OECD Healthcare Quality Project
- OECD Improving Value in Healthcare: Measuring Quality
- Conceptual Model for National Healthcare Quality Indicator System in Norway
- Denmark Quality Indicators
- UK NICE standards – Selecting and Prioritizing Quality Standard Topics
- Australia's – Indicators used Nationally to Report on Healthcare, 2013
- European Commission Healthcare Quality Indicators
- Consumer-Purchaser Disclosure Project – Ten criteria for meaningful and usable measures of performance

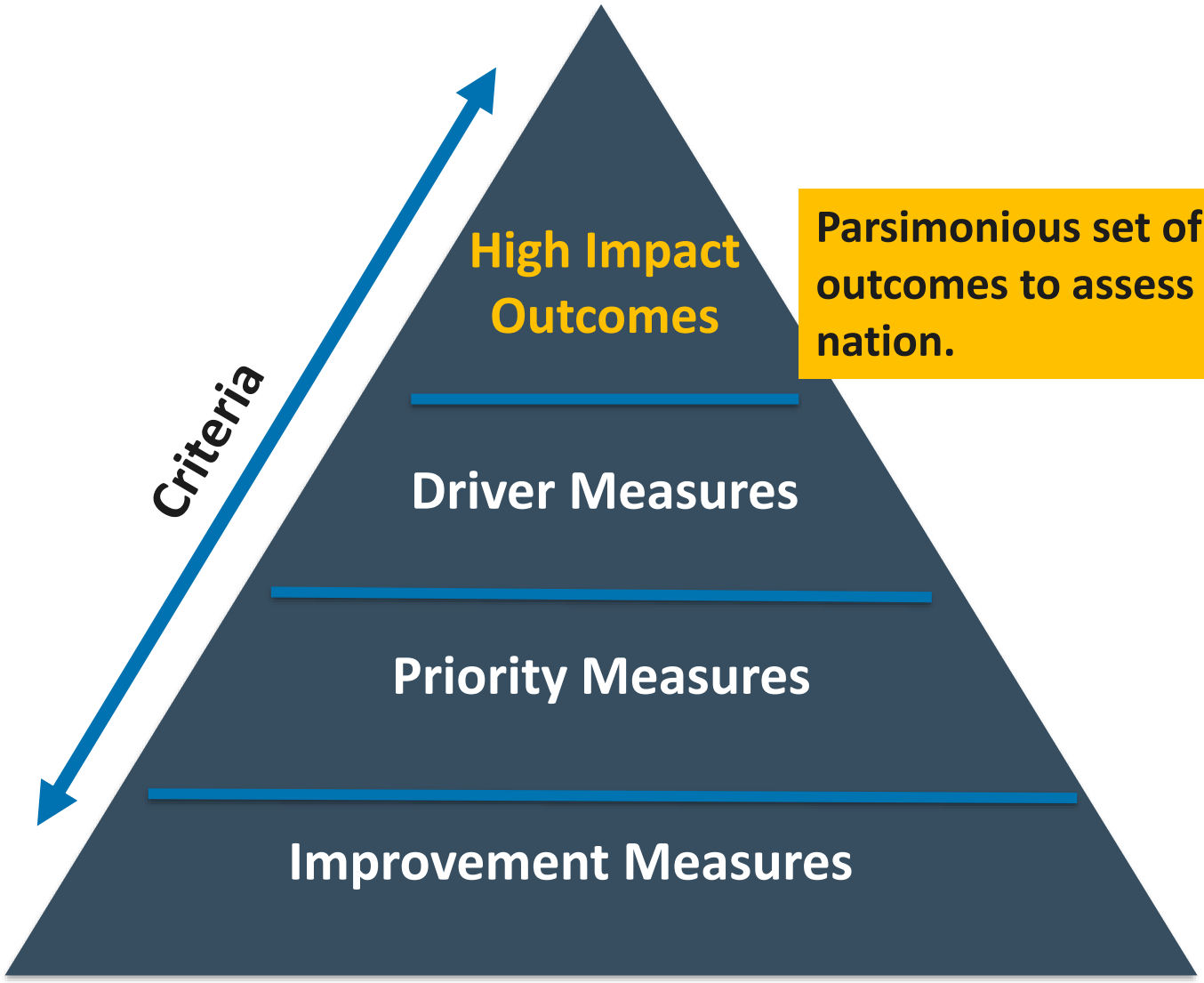
NQF Prioritization Criteria

Criterion	Description
Outcome-focused	Preference for outcome measures and measures with strong link to improved outcomes and costs
Improvable and actionable	Preference for actionable measures with demonstrated need for improvement and evidence-based strategies for doing so
Meaningful to patients and caregivers	Preference for person-centered measures with meaningful and understandable results for patients and caregivers
Support systemic and integrated view of care	Preference for measures that reflect care that spans settings, providers, and time to ensure that care is improving within and across systems of care

Hierarchical Framework



Hierarchical Framework

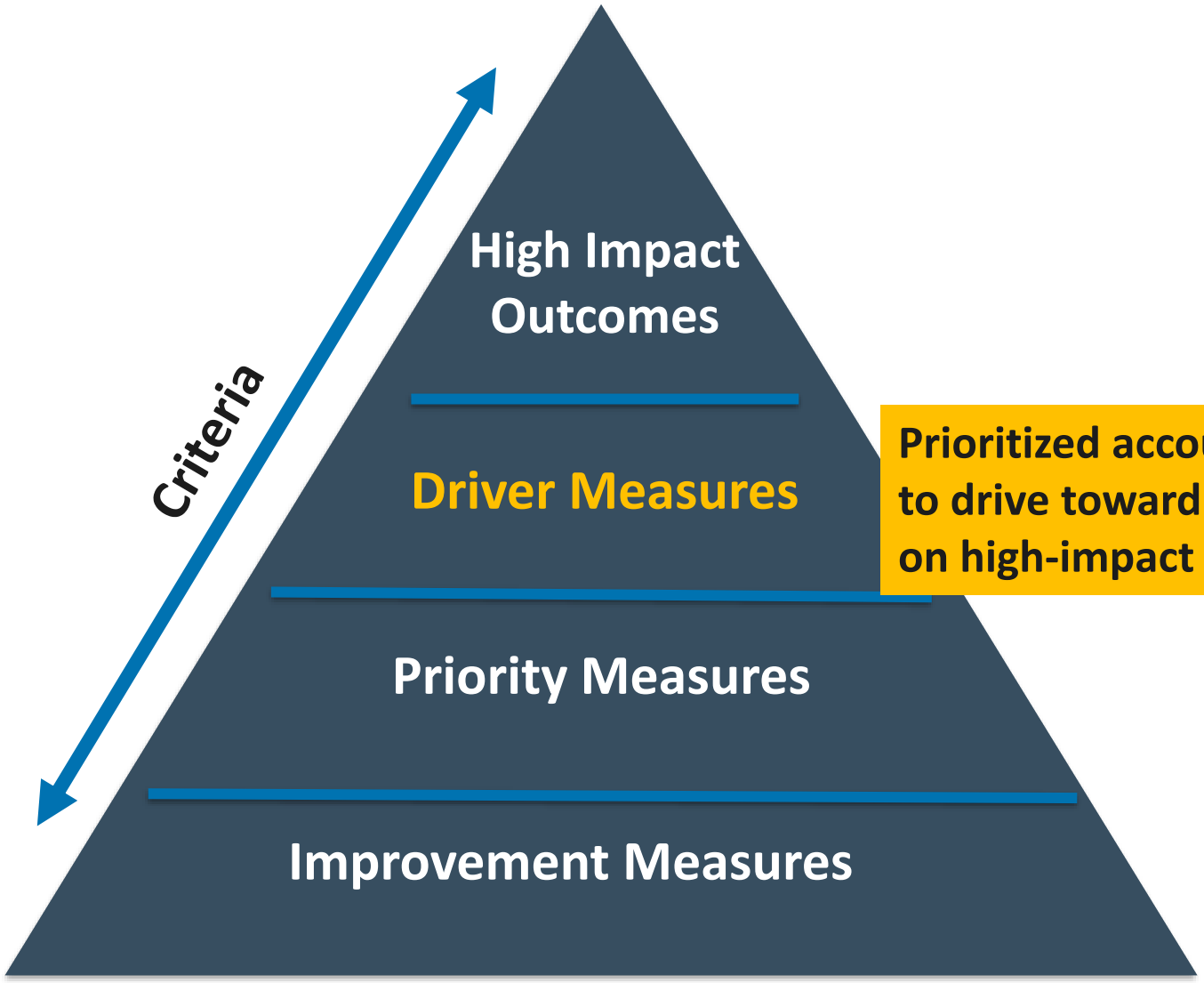


Parsimonious set of high-impact outcomes to assess progress as a nation.

High-Impact Outcomes

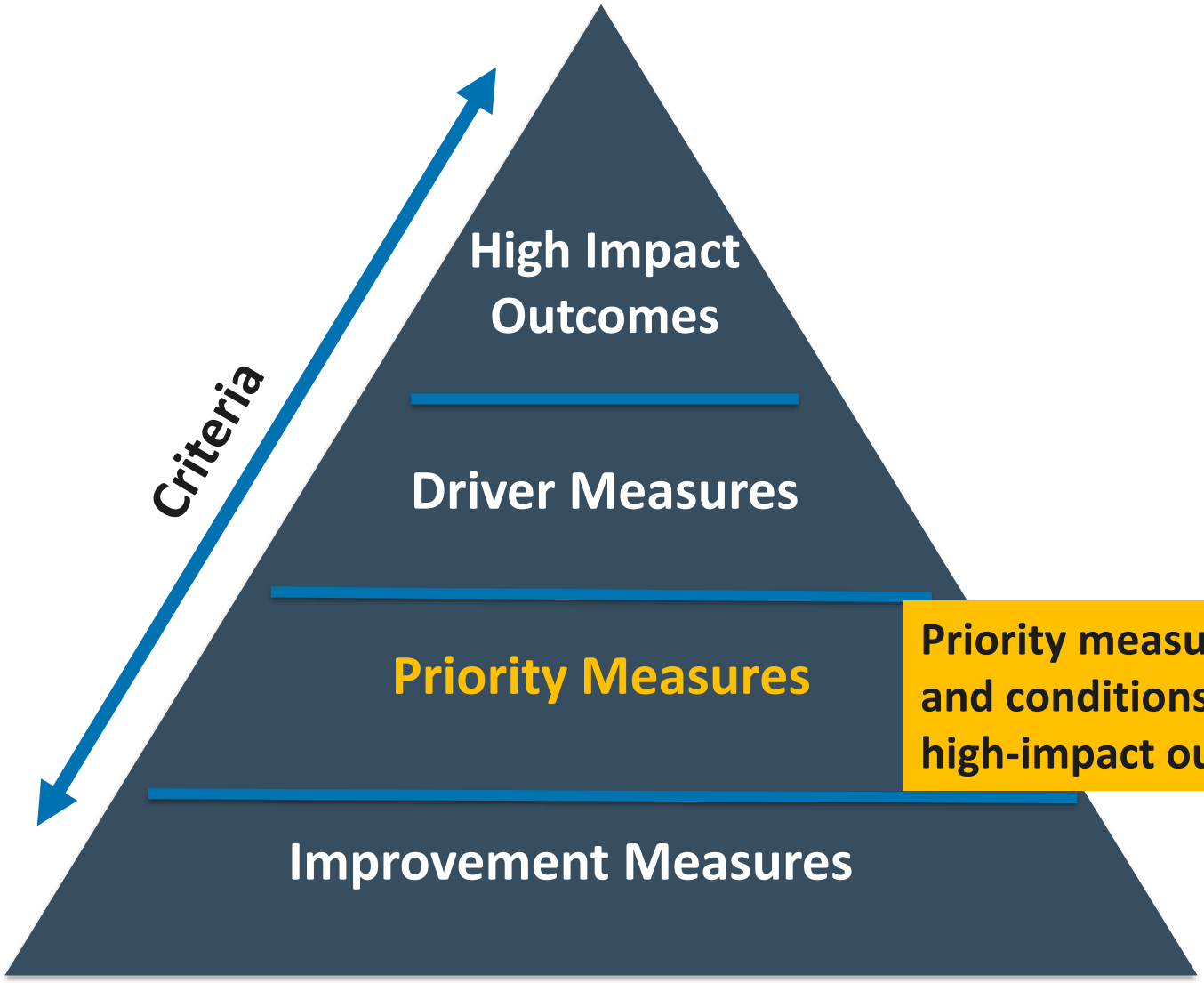
High Impact Outcomes
Functional status/well-being
Patient experience (including care coordination, shared decisionmaking)
Preventable harm/complications
Prevention/healthy behaviors
Total cost/high-value care
Access to needed care
Equity of care

Hierarchical Framework



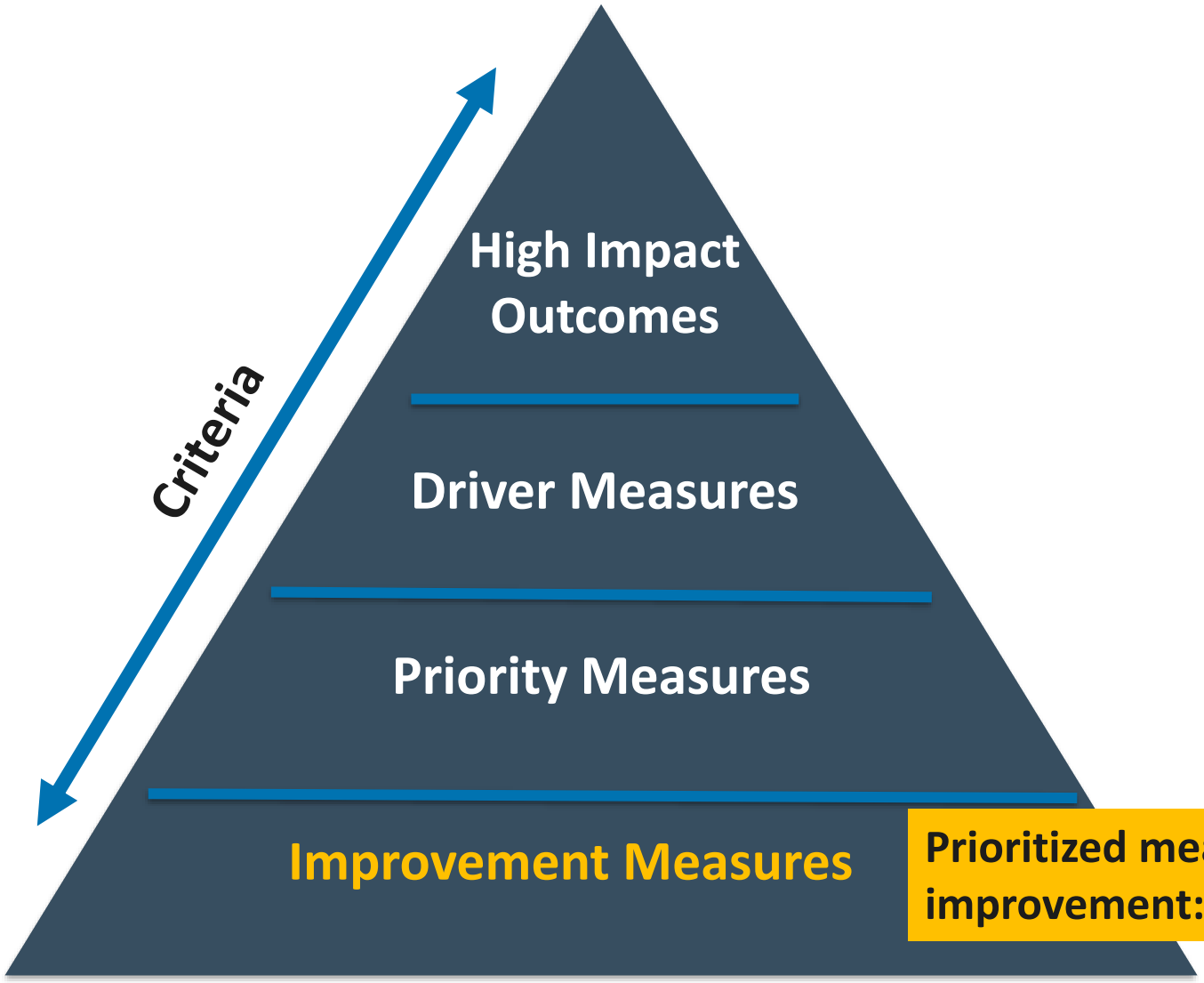
Prioritized accountability measures to drive toward higher performance on high-impact outcomes.

Hierarchical Framework



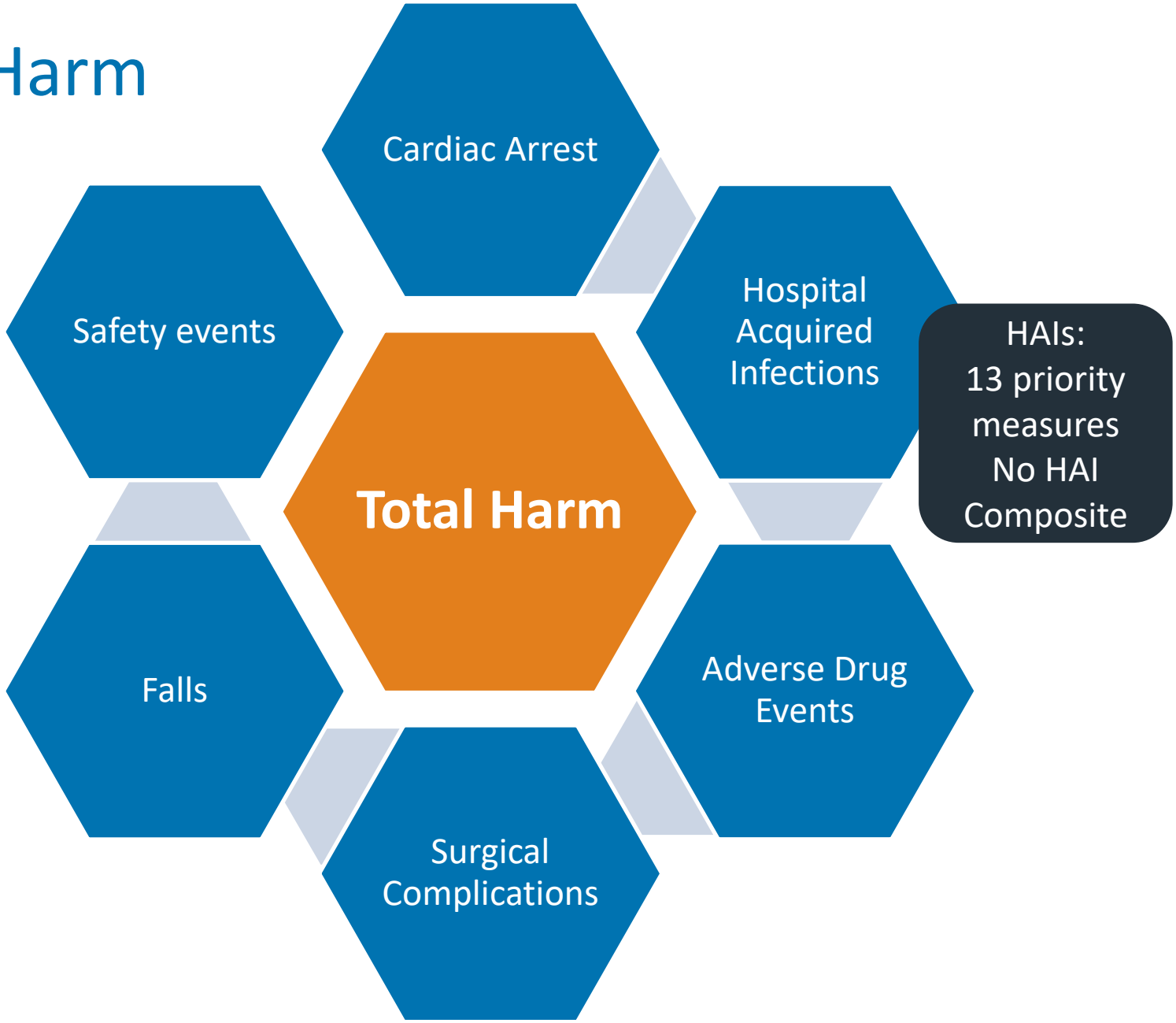
Priority measures in specific settings and conditions that contribute to high-impact outcomes.

Hierarchical Framework

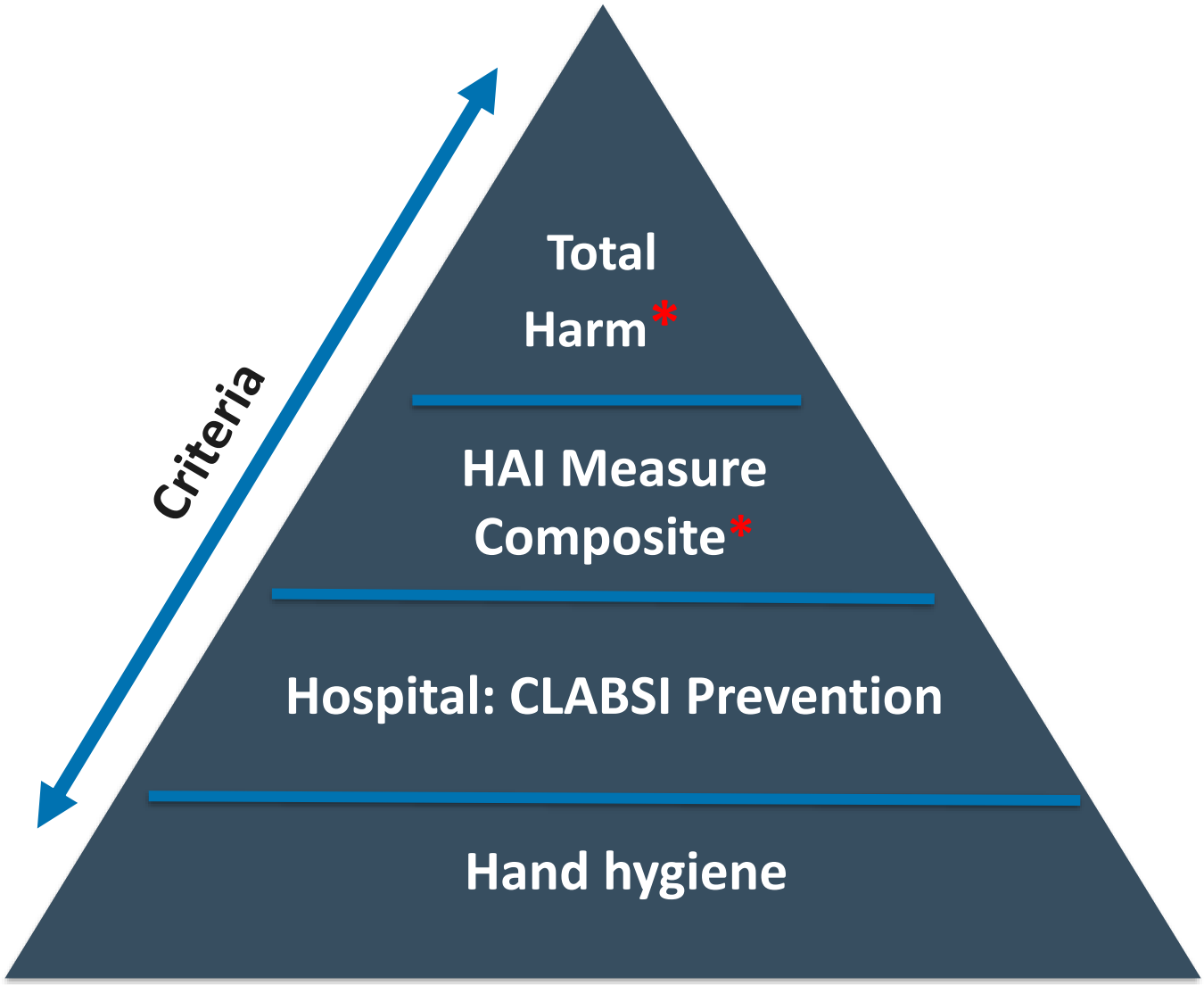


Prioritized measures to drive improvement: standardize & share

Total Harm


Cardiac Arrest
Safety events
Hospital Acquired Infections
HAIs:
13 priority measures
No HAI Composite
Total Harm
Falls
Adverse Drug Events
Surgical Complications

Example: Total Harm



* Gap

Prioritization: Next Steps

- Develop approach to assess the attributable effect of potential measures that can drive toward improved performance on the high-impact outcome
- NQF will build the prioritization approach for measures and gaps into future endorsement and selection work
- NQF will explore potential partnerships to share and standardize prioritized improvement measures

Graduated Measurement

Persistent gaps in measurement

- The pipeline of innovative measures to support healthcare system transformation is limited
- Current measures cannot fully support shift to alternative payment models and population focus
- Persistent barriers in the development, testing and use of innovative measures:
  - Access to funding
  - Access to test beds
  - Use of innovative measures in accountability programs

How can NQF support measure innovation?

- Consensus process designed for fully developed measures intended for accountability
- New pathways could bring measures to NQF along the measure development and testing cycle
- Provide access to multistakeholder input and NQF guidance *throughout* the measure development lifecycle
- More flexible approach should encourage access to:
  - Earlier availability of measurement expertise
  - Access to multistakeholder input
  - Direct input and approval from NQF standing committees at different stages of development and testing

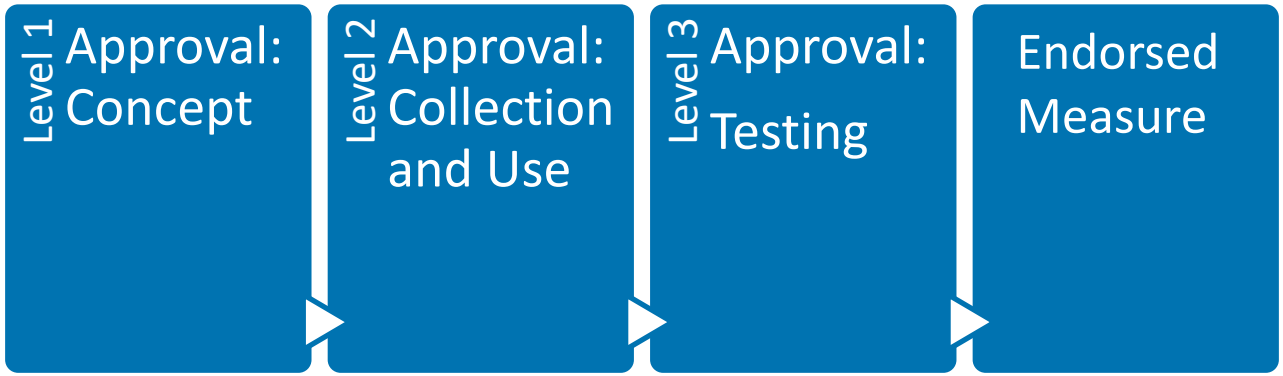
Prior Experience with Flexible Submission

- Time limited endorsement
- Two-stage endorsement pilot
- eMeasures for Trial Use

Graduated Measurement Approach

- Build on prior models and develop new agile approach for measure submission
- Measure developers could enter the process at any stage to meet the needs of their current state of development
- Earlier submission for multistakeholder input and expert input could encourage use and feedback prior to incorporation into accountability programs
- More agile process should allow NQF committees to approve measures at earlier stages of development for use and testing
- NQF Measure Incubator can be accessed at any stage of development

Draft Graduated Measurement Levels



Level 1: Concept Approval

- Potential options:
  - Multistakeholder feedback through NQF commenting
  - NQF technical assistance
- No standing committee engagement
- Measure implementation: None

Level 2: Approval for Collection and Use

- Standing committee review:
  - *Evidence*
  - *Usability*
  - *Precision of draft specifications*
  - *Preliminary feasibility assessment based on data source*
- Measure implementation - collect data in the field and iteratively improve measure prior to testing

Level 3: Approval for Testing

- Standing committee review:
  - *Evidence*
  - *Usability*
  - *Feasibility*
  - *Performance gap*
  - *Precise specifications*
  - *Face validity*
- Measure implementation - share data and analytics – useful for improvement and benchmarking; not ready for payment/public reporting

Level 4: Endorsement

- Standing committee review: as above, with formal testing for reliability and validity
- Consider additional requirements for endorsement (e.g., measure score level testing, feedback)
- Measure implementation – use for accountability (payment/public reporting)

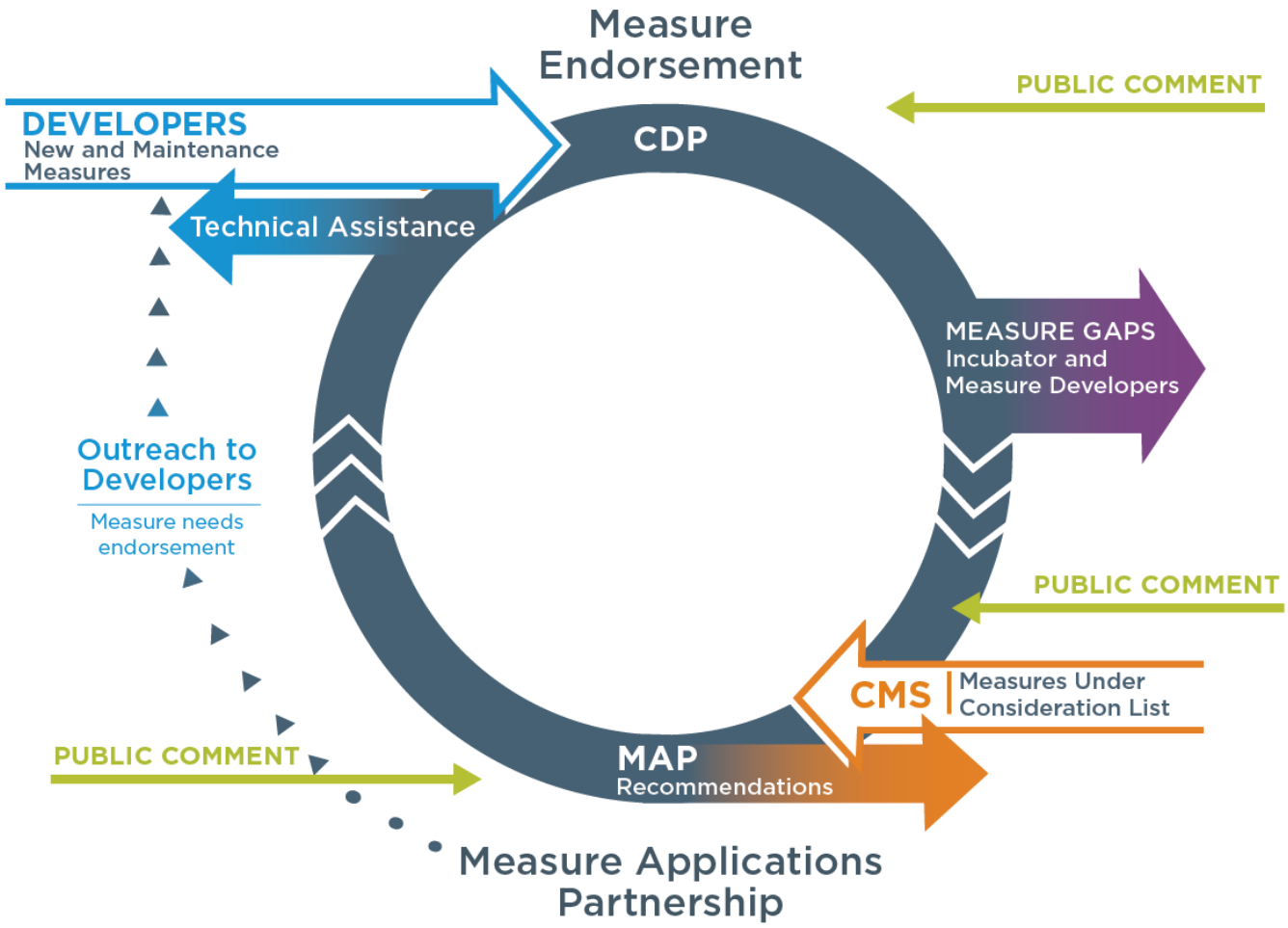
NATIONAL QUALITY FORUM

Feedback on Measures in Use: What Can We Learn?

John Bernot, MD

NQF Strategic Vision

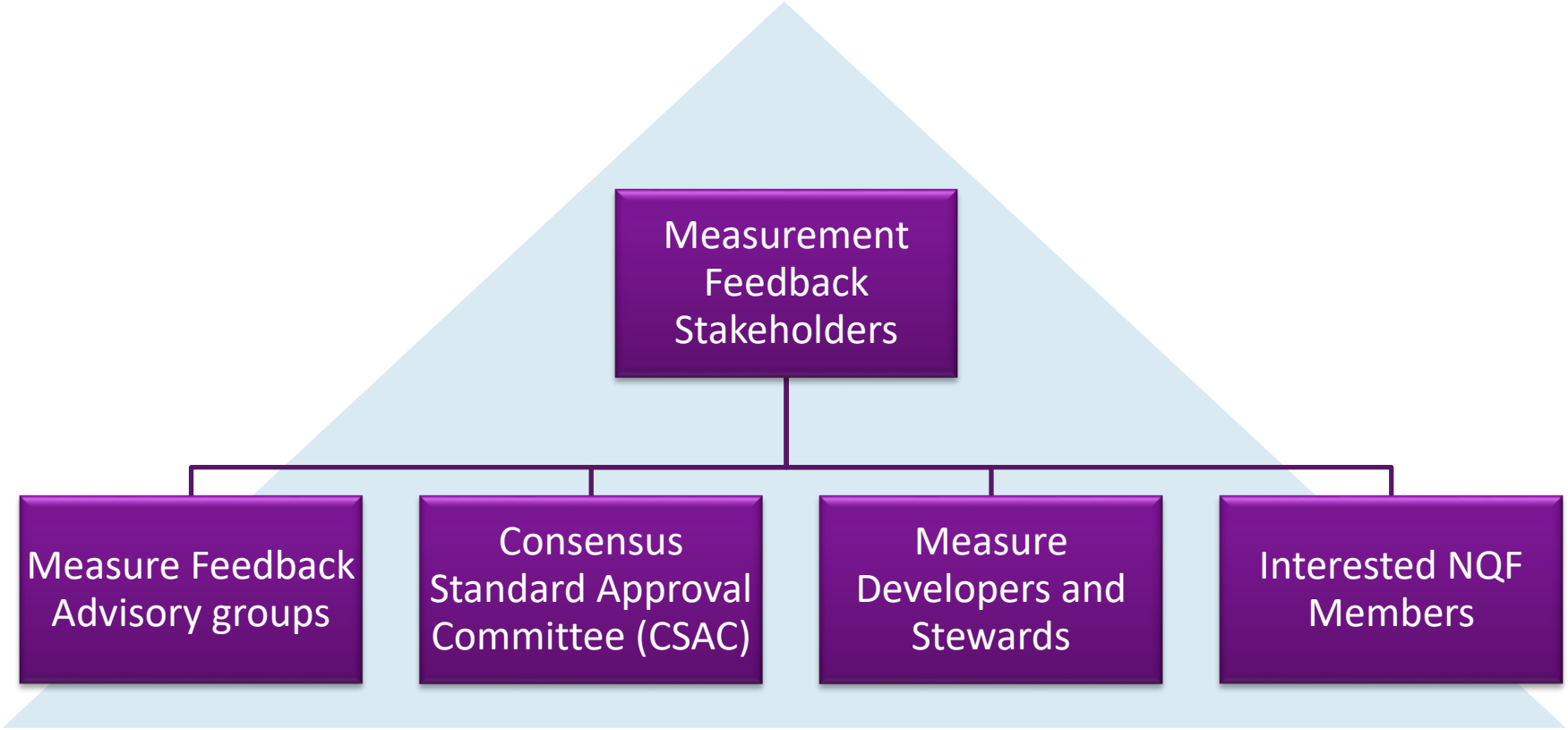
Current Structure: Collecting and Integrating Feedback



Objectives

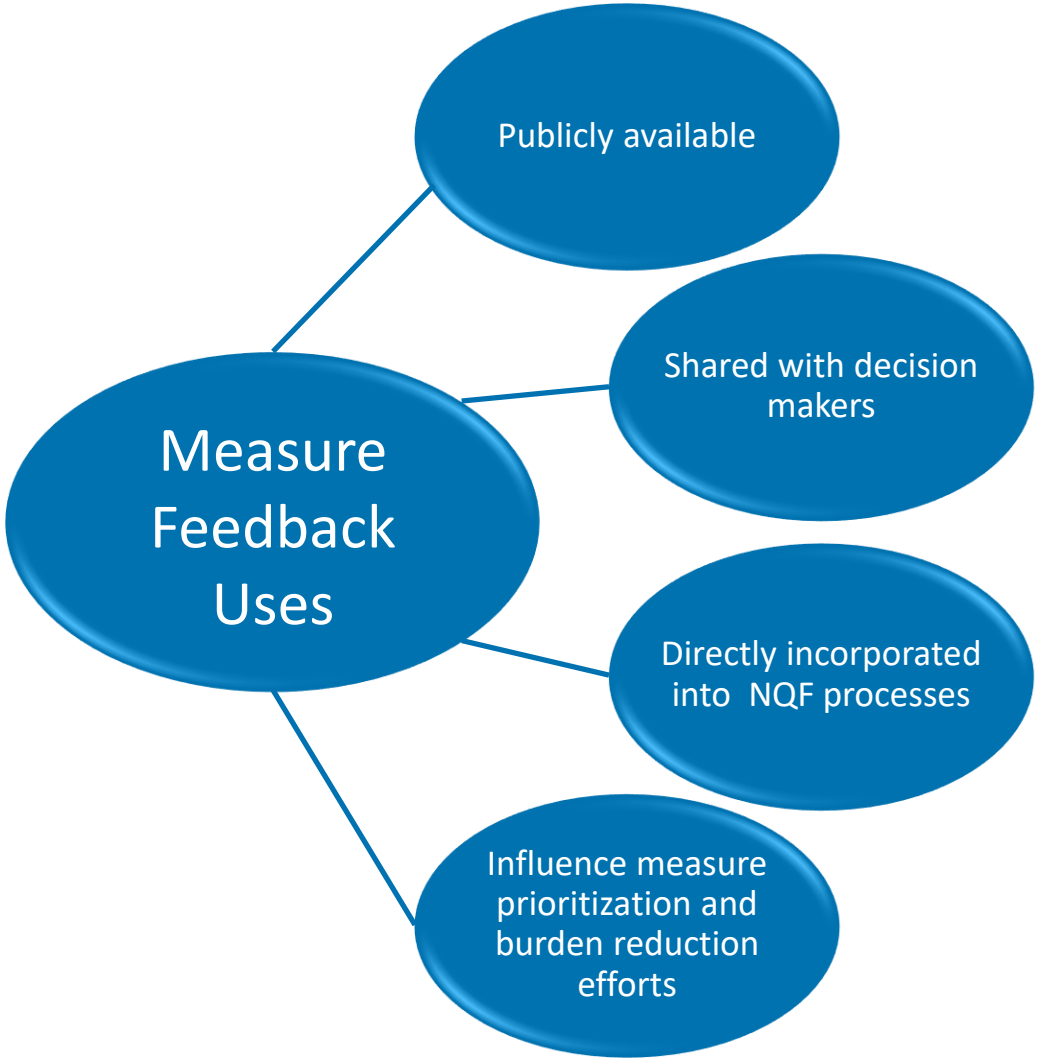
- Learn from the field about experiences with measures
- Enhance and expand feedback on the implementation and impact of measures
- Share learnings broadly within and outside of NQF
- Inform measure prioritization and reduction

Identifying Stakeholder Priorities



Collecting Measure Feedback

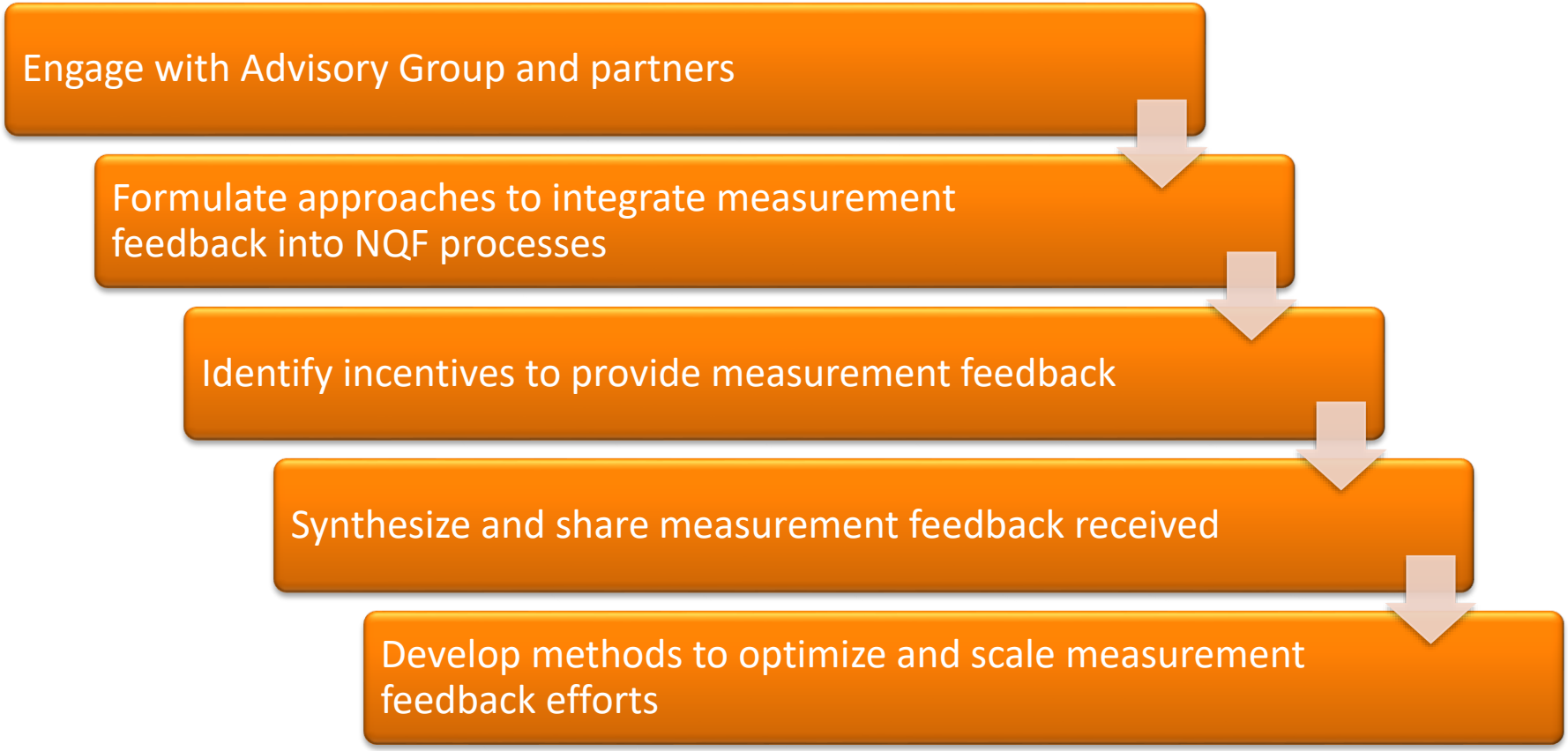
Measure Feedback Uses



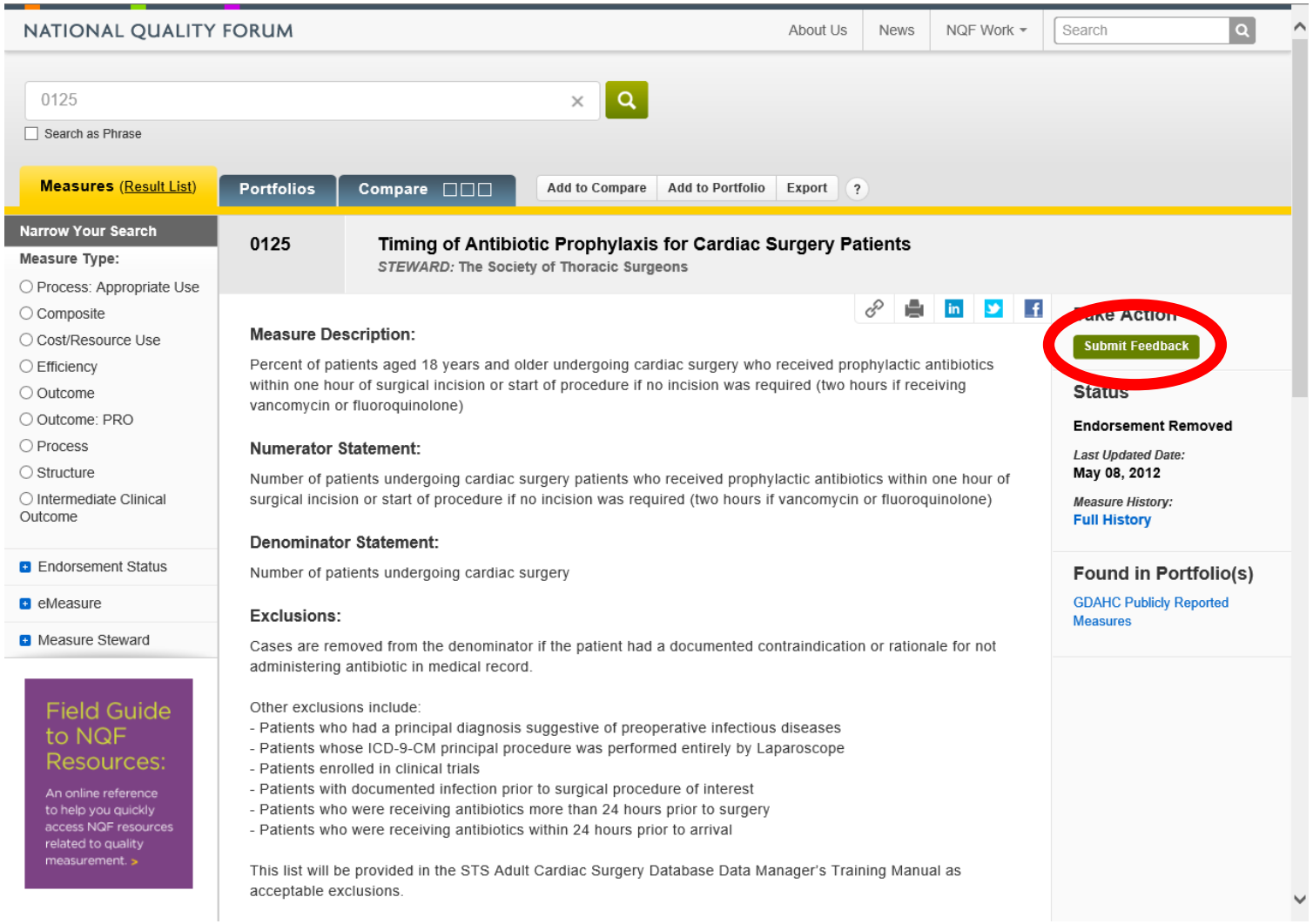
Decision Makers

- Measure Applications Partnership (MAP)
- Measure stewards and developers
- NQF endorsement committees
- NQF Measure Incubator
- Key stakeholders

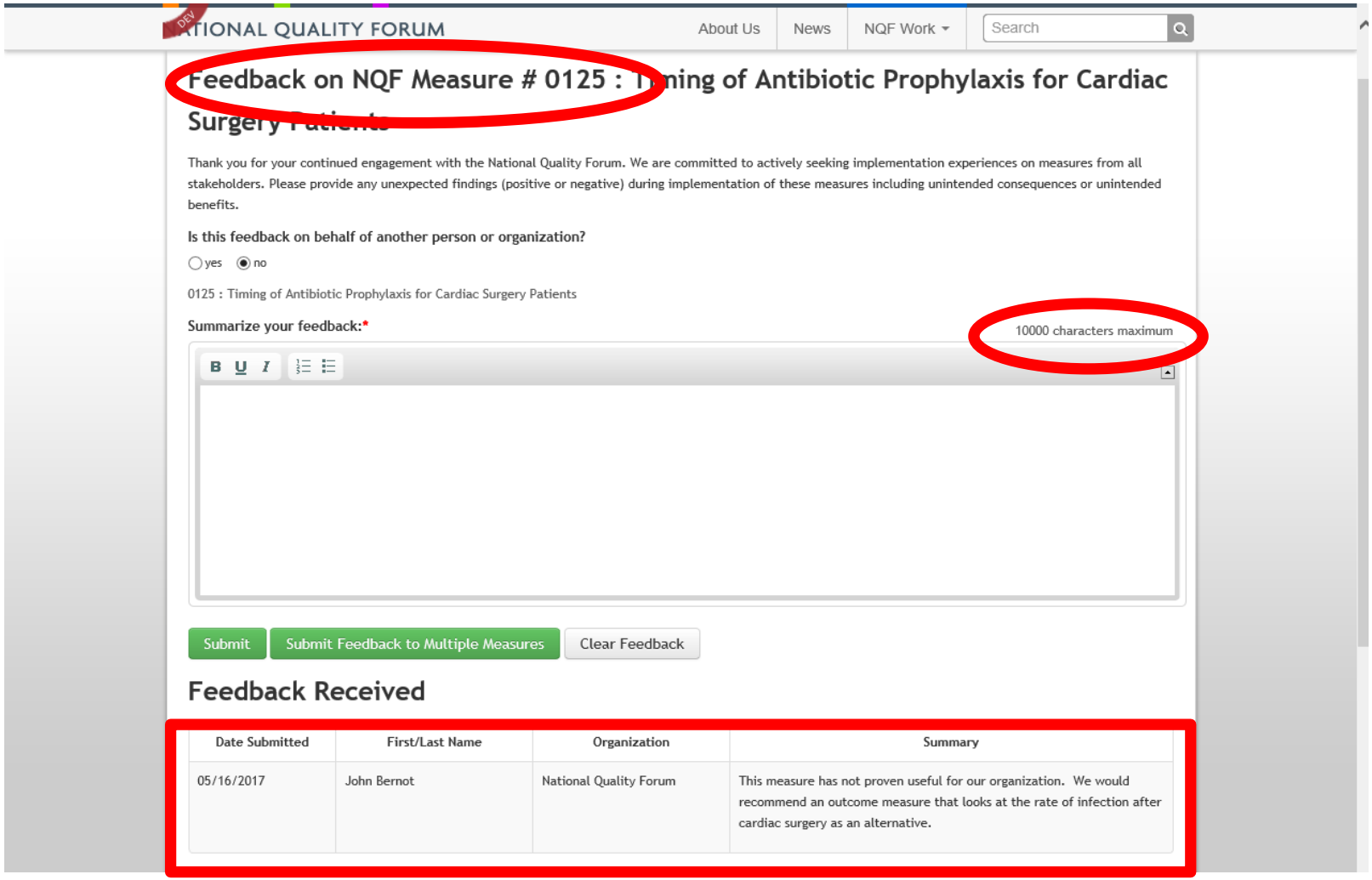
Next Steps


Engage with Advisory Group and partners
Formulate approaches to integrate measurement feedback into NQF processes
Identify incentives to provide measurement feedback
Synthesize and share measurement feedback received
Develop methods to optimize and scale measurement feedback efforts

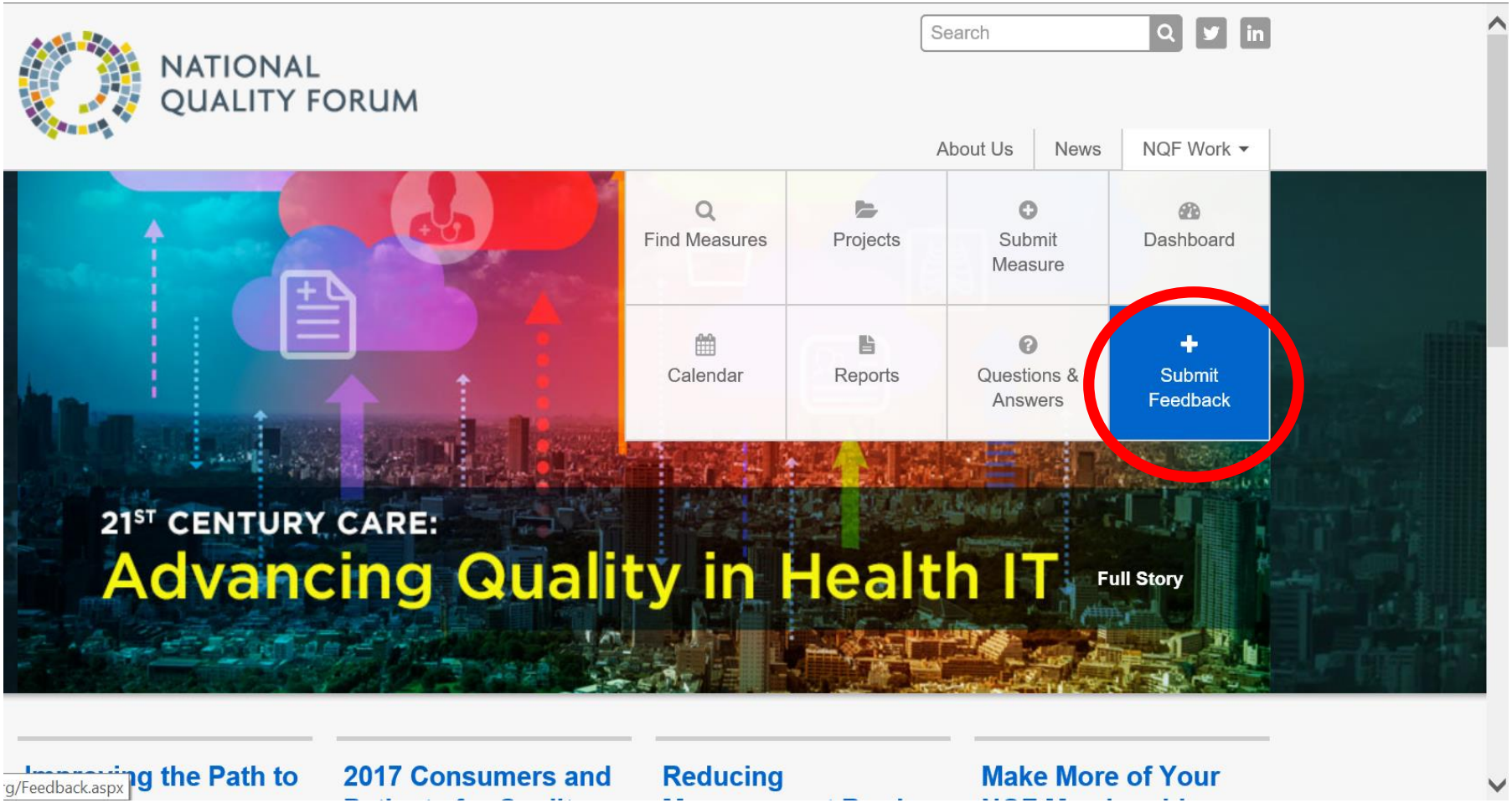
Demonstration



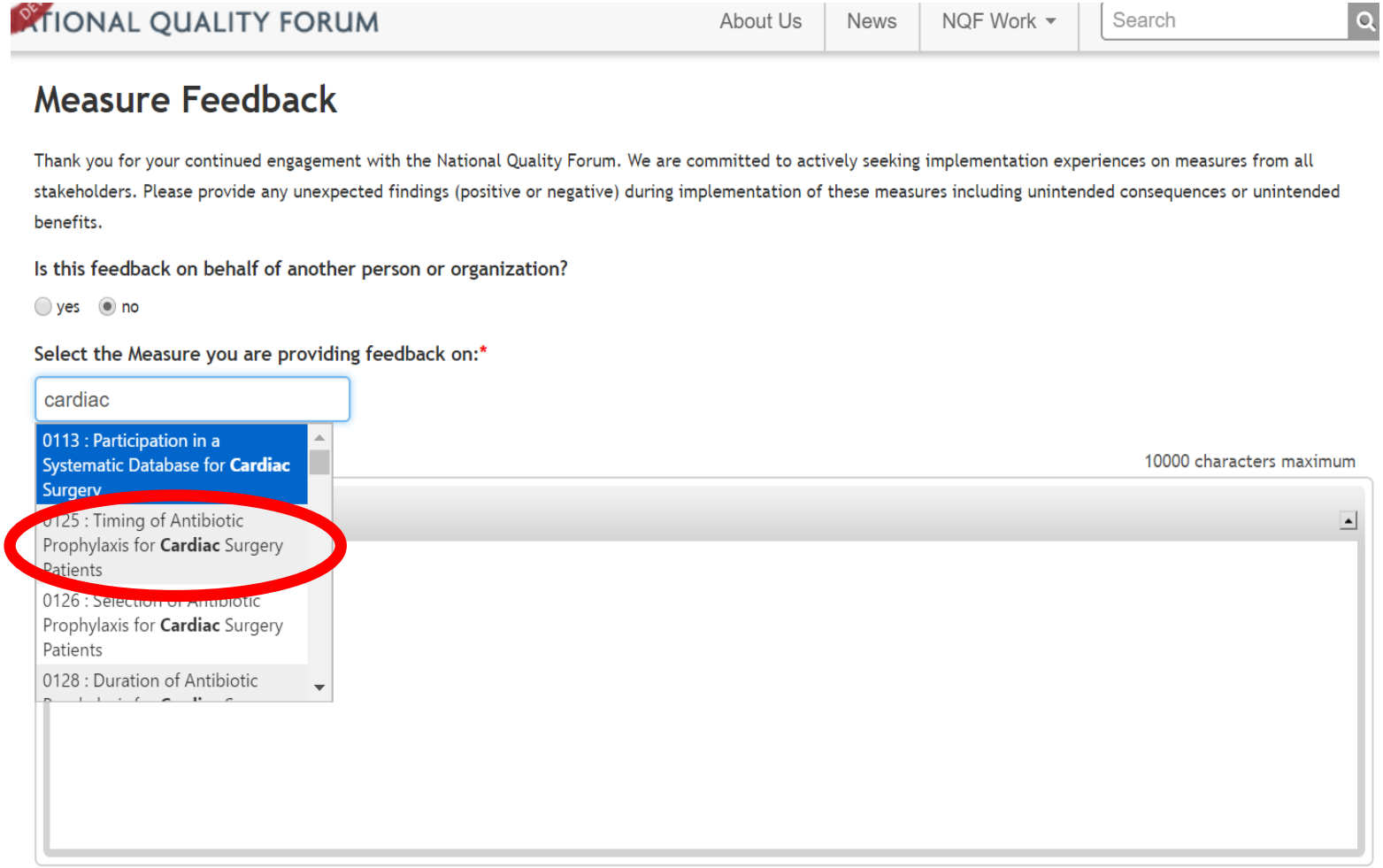
Demonstration



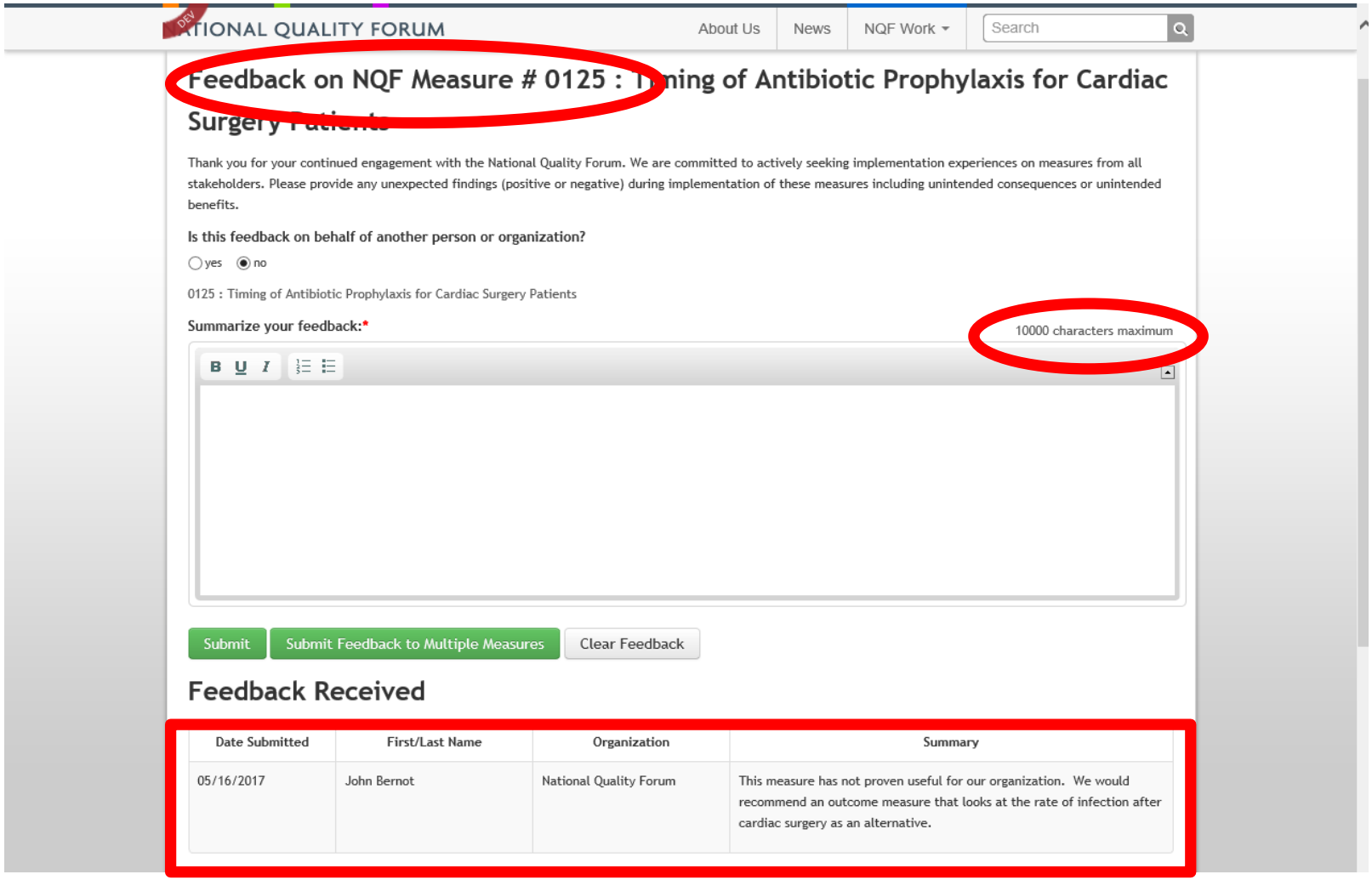
Demonstration



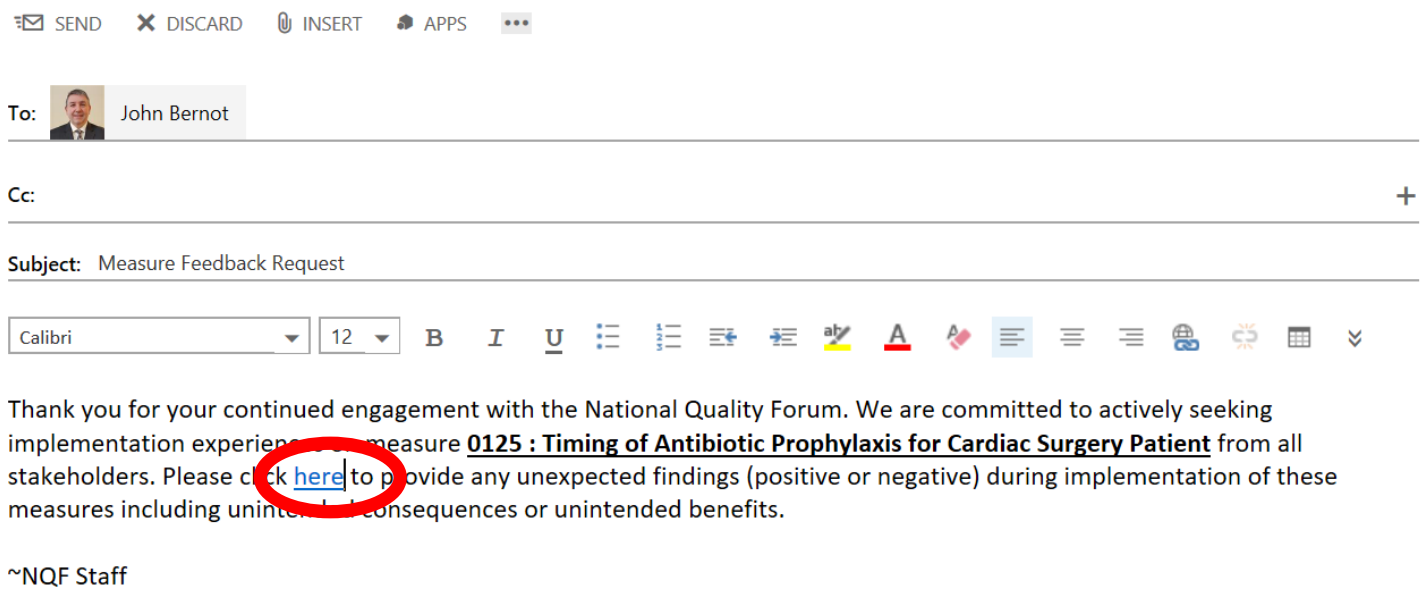
Demonstration



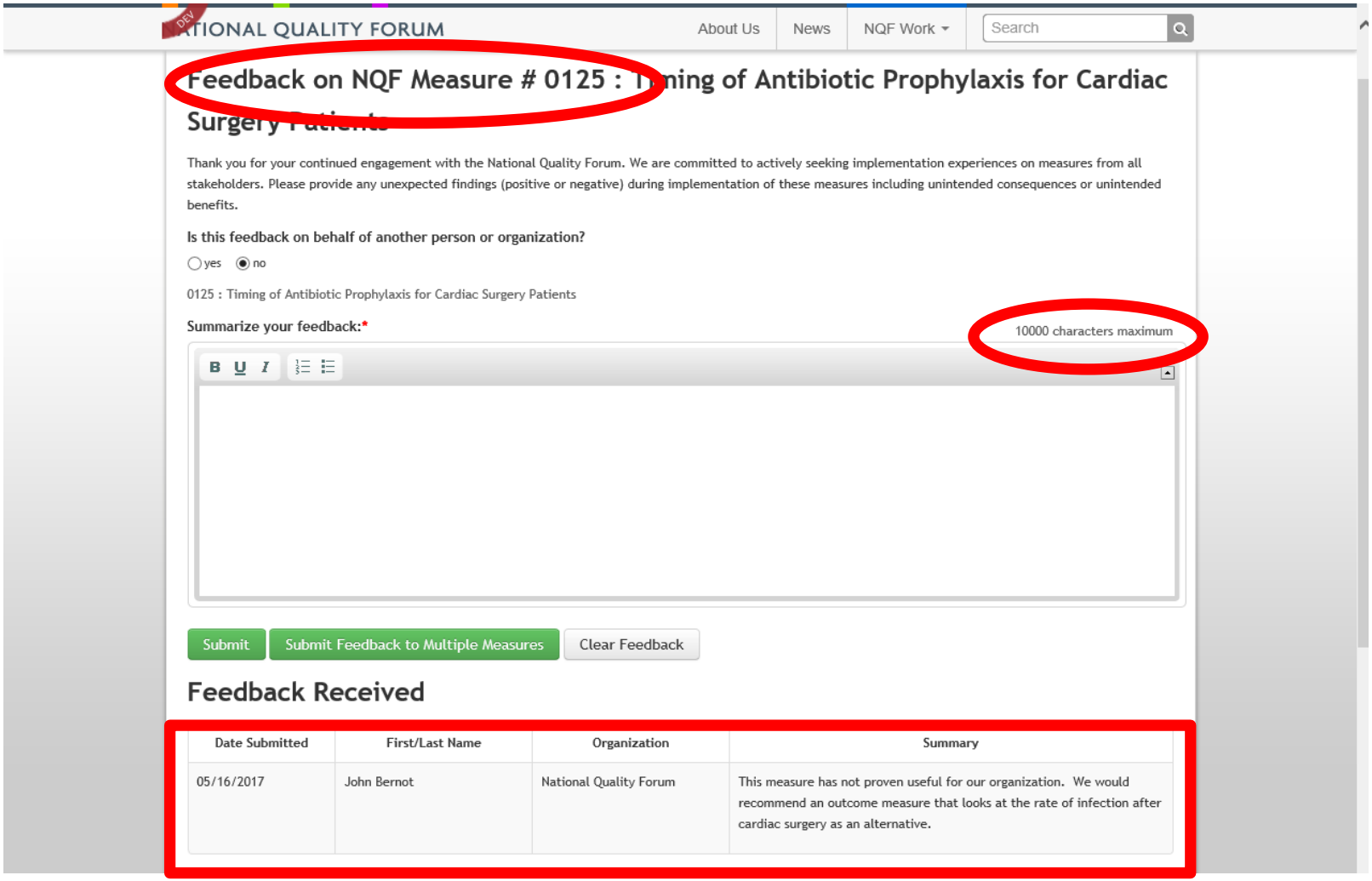
Demonstration



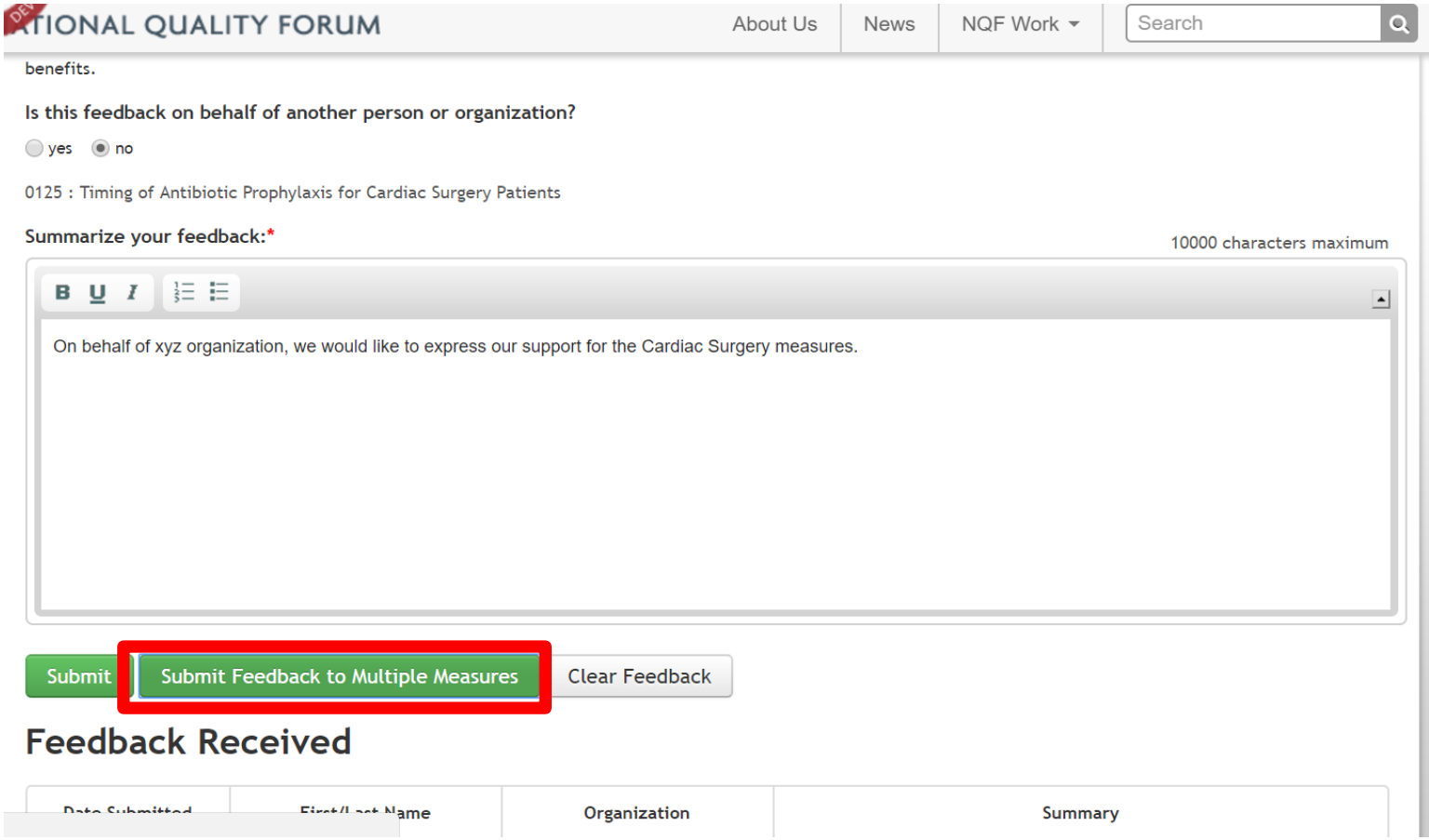
Demonstration



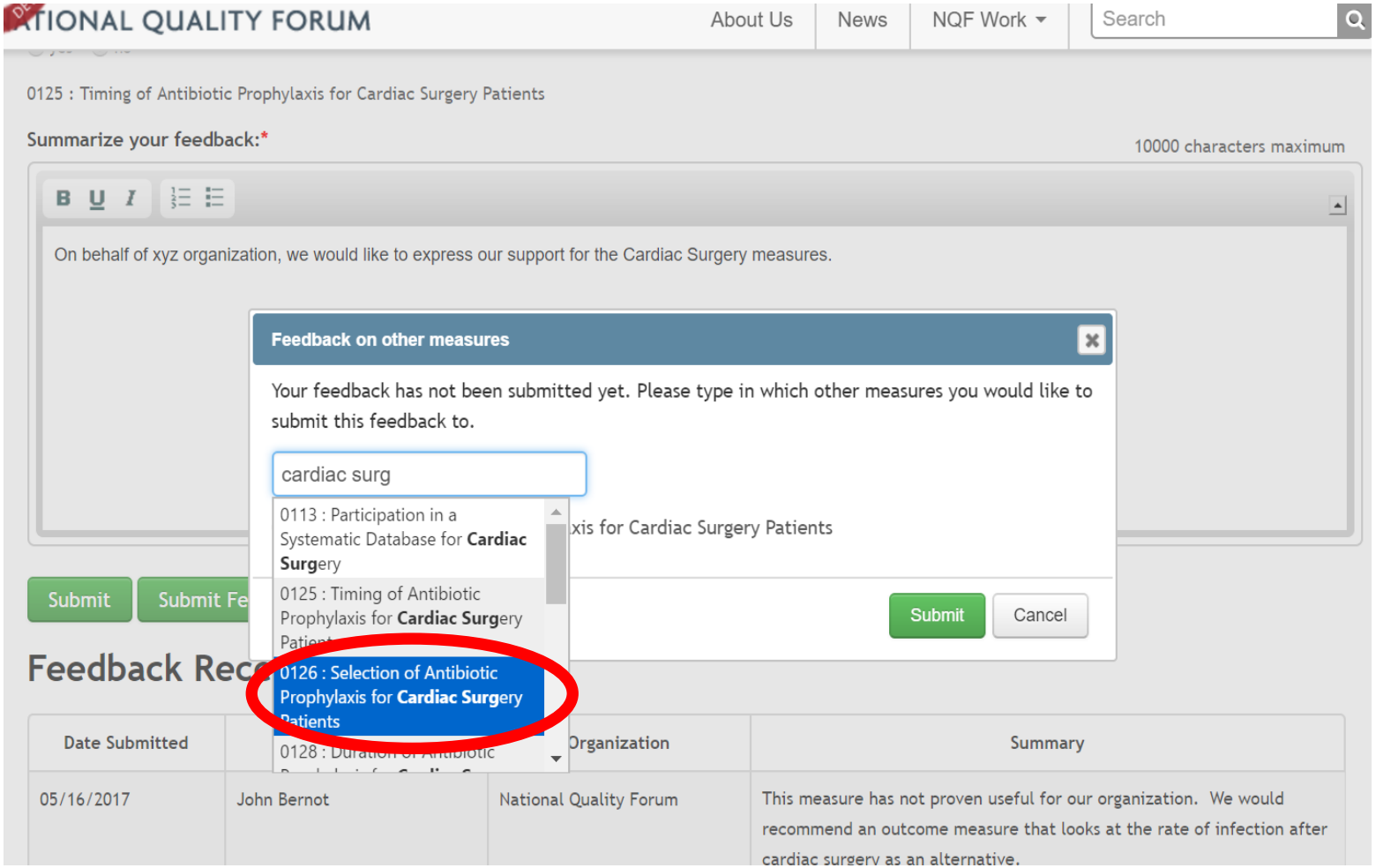
Demonstration



Demonstration



Demonstration



Questions?

Break

NATIONAL QUALITY FORUM

Measure Incubator™ Update
A Focus on Patient-Centric Performance Measures

Kathleen Giblin, Senior Vice President, Quality Innovation, NQF

Lyn Paget, Consultant, PatientsLikeMe

May 17, 2017

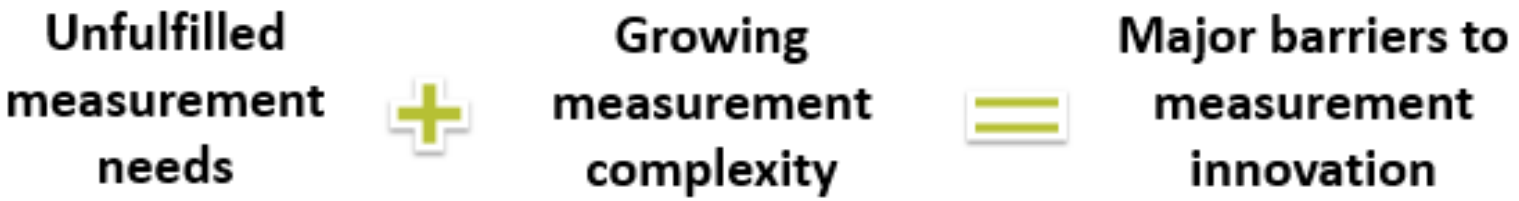
Measure Incubator Update

Strategic Direction

Improve healthcare quality, safety, and affordability

NQF Measure Incubator

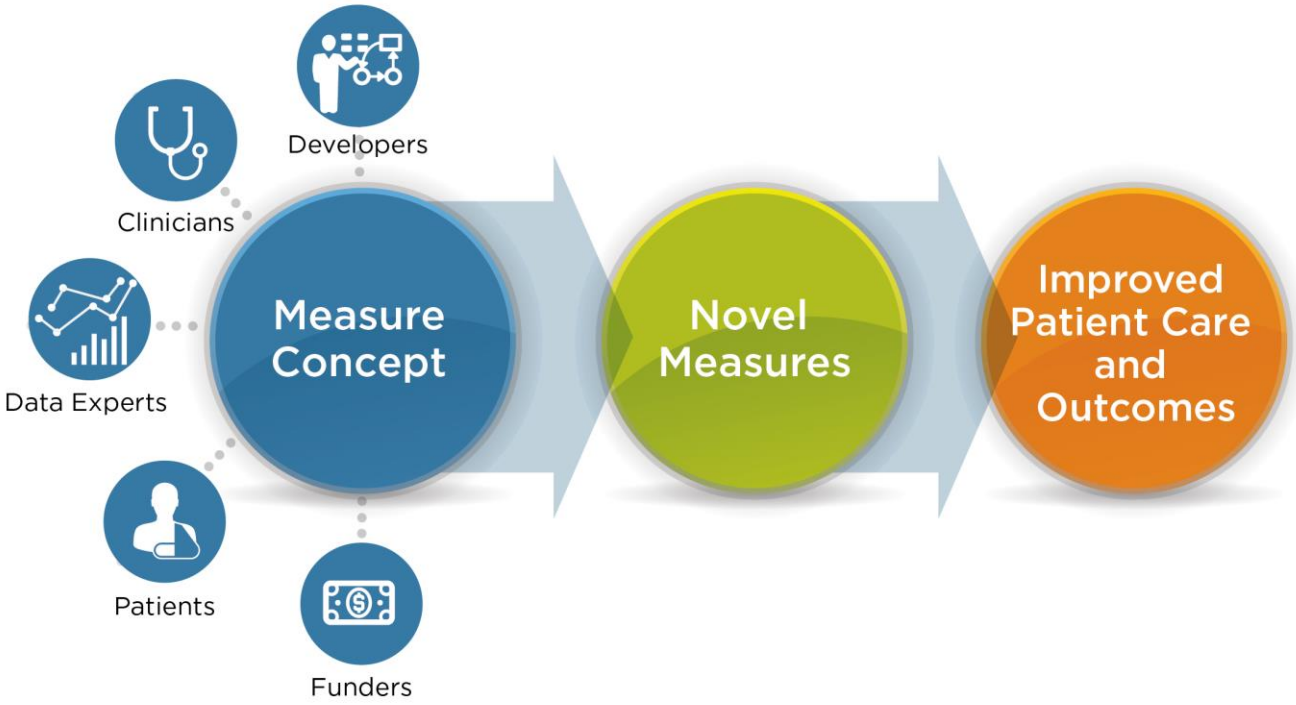
The Drive for Meaningful Quality Measures



- Current measure development is slow, costly and rigid
- Most existing measures are built from administrative claims/paper medical records alone
- Highly complex specification and testing processes
- Need for fewer, high-impact measures that assess:
  - *Outcomes and composites*
  - *System-level performance*
  - *Patient-reported experience and outcomes*

NQF Measure Incubator

Getting to Quality Measures that Matter



NQF Measure Incubator

The Learning Collaborative

- **Purpose:** identify and share best practices around tough issues of measure development and to creatively collaborate on identifying solutions

- **Includes** stakeholders interested in measurement, such as measure developers, researchers, data entities, purchasers, patient organizations, and clinician groups

Measure Incubator Learning Collaborative

2016 Innovation Challenge

Five winners proposed novel methodological approaches to improving healthcare quality measurement

- **Charlotta Lindvall, Dana-Farber Cancer Institute:** Proposes using natural language processing to develop quality measures in palliative surgery using electronic health record (EHR) data
- **S. Mani Marashi, Henry Ford Health System:** Describes a successful two-year pilot to report hospital-acquired venous thromboembolism events in real-time using data from EHRs, rather than claims
- **Robert Philips, American Board of Family Medicine:** Proposes using a new data registry open to all primary care physicians to identify and develop efforts to improve clinical practice and quality measures
- **Ellen Shultz & Michelle Langer, American Institutes for Research:** Suggest using "bookmarking," a method widely used in educational testing, to score and classify patient-reported outcome (PRO) measures and address this critically important measure gap area
- **Tracy Spinks, MD Anderson Cancer Center:** Outlines a new, streamlined, standardized approach to implementing PRO measure sets in EHRs

NQF Measure Incubator

Current Projects

- Multiple Sclerosis (MS) PRO-PM
- Chronic Obstructive Pulmonary Disease (COPD) PRO-PM
- Rheumatoid Arthritis PRO-PM
- Collaboration with PatientsLikeMe informs incubation of PRO-PMs
- Obesity PRO-PM
- Pain Management in Acute Care Settings
- Advanced Illness Care
- Oncology

Patient-Reported Outcomes a Main Focus of Measure Incubator Projects

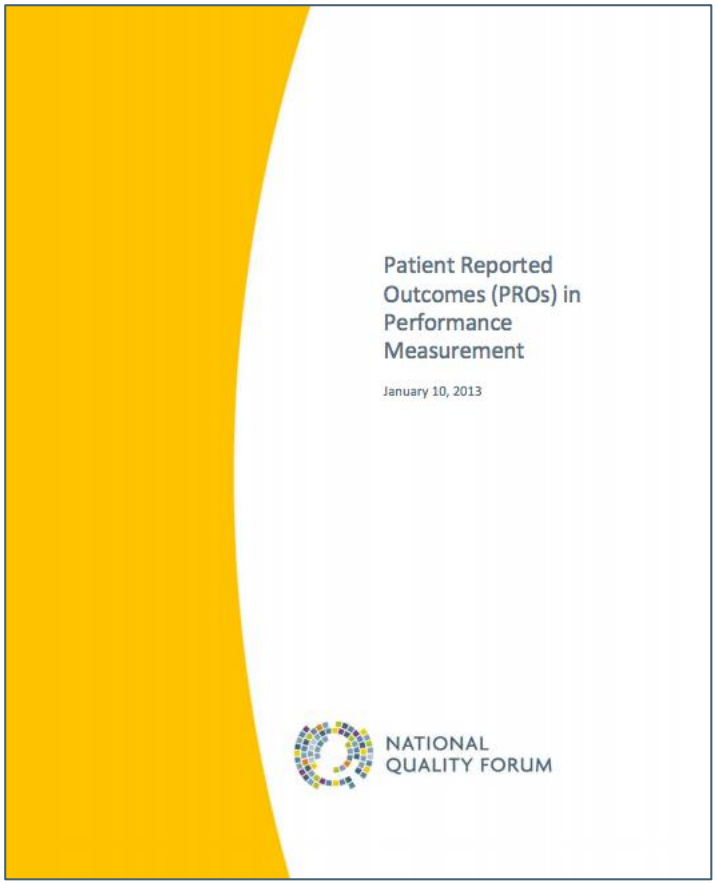
Building Patient-Centric Performance Measures

Background

- ✓ Patients are increasingly being viewed as the essential data source to help evaluate, improve and guide payments for health care.
- ✓ Patient prioritized measurement is front and center in the 2016 CMS quality measure development plan.
- ✓ Patient-reported outcomes (PROs) - information on the outcomes of health care obtained directly from patients - have become the most prominently recommended shift in the paradigm of quality improvement.

Project Overview

- 2015 Robert Wood Johnson Foundation funded PatientsLikeMe in partnership with NQF to demonstrate a model that incorporates patient priorities into performance measurement.
- Building on the 2013 NQF recommendations and guiding principles for PRO Performance Measurement.



Robert Wood Johnson Foundation

Project Components

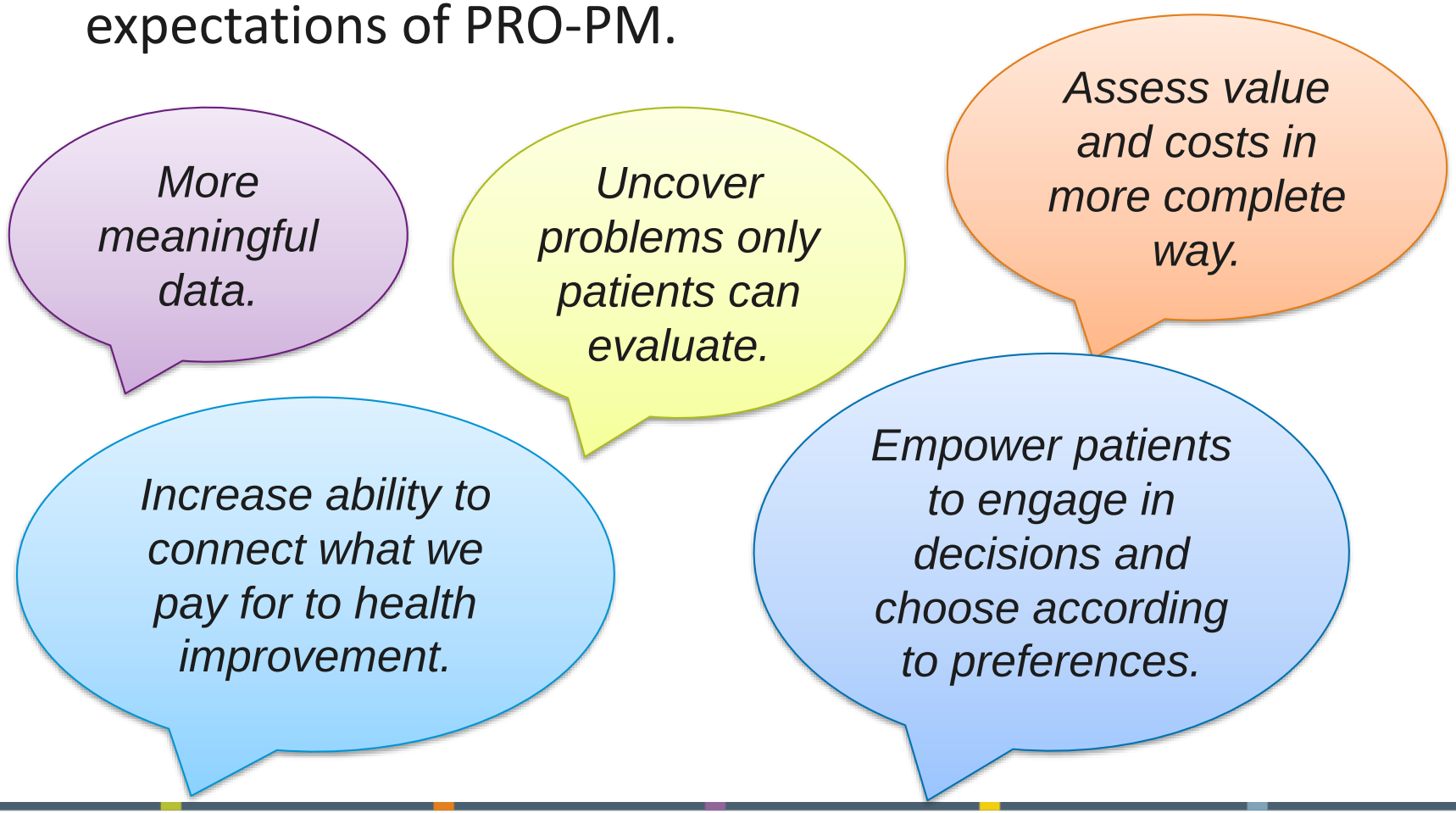
1. Explore needs of measurement field;

2. Develop a conceptual model of good health care from the patient's perspective; and

3. Demonstrate methods to incorporate patient experience data into measurement.

Measurement Field Needs

- Exploratory scan including 26 multi-stakeholder interviews to better understand the requirements and expectations of PRO-PM.



Implementation Facilitators

- ✓ Leverage technology to identify patients, minimize burden, and support shared decision-making.
- ✓ Talk with patients about measures and how they are used.
- ✓ Neutralize workflow impact by offloading other tasks.
- ✓ Institute incentives and reimbursement models to advance PROM use.
- ✓ Risk adjust PROMs for fair provider comparison and to restrict cherry picking of patients.

Implementation Facilitators

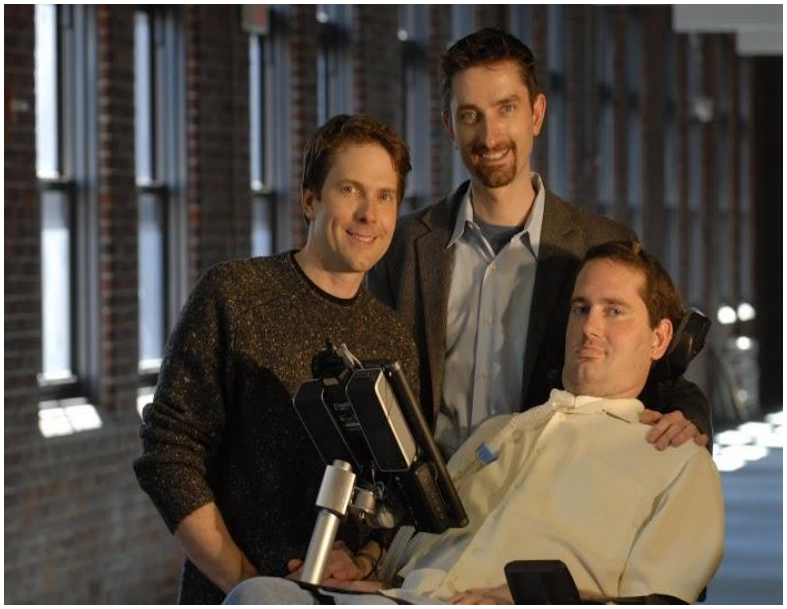
- ✓ Choose PROMs calibrated for individual and population level analysis.
- ✓ Select PROMs that match the needs and goals of patients and clinicians (not researchers).
- ✓ Offer multiple modalities for patient data input.
- ✓ Provide interpretation of data that is actionable for patients and clinicians.
- ✓ Translate and adapt PROMs for diverse patient populations.

Illustration of Value


patientslikeme®

About PatientsLikeMe

- Founded in 2004 as a direct response to a family's experience with chronic disease (ALS)
- Open, patient facing research based community
- Opened to any condition in 2011
- Deep patient experience data ~40 chronic life-changing conditions
- Free to join. No advertising



Patients	Data	Insights
• 500,000+ members • 2,700+ conditions • 100% volunteer	• 38+ million structured data points • 4+ million free-text posts • 15+ PROs • Patient-generated taxonomy (mapped to clinical classification systems)	• 90+ publications • Staff: research, clinical, science, technology

PLM Data Domains

- Basic Information (age, sex, etc.)
- Diseases (early signs, diagnosis status, etc.)
- General & Specific Symptoms (onset, severity status, etc.)
- Treatments & Side Effects (Rx, OTC, Supp., non-drug, etc.)
- Quality of Life & Behavior Status (all patients, some disease specific)
- Outcome Measures of Disease (disease dependent)
- Patient-generated narrative text, wearable and sensor data

- Engagement
- Data Integrity
- Standards *
- Evidence
- Knowledge
- Empowerment

* Patient data mapped to ICD-10, ICF, SNOMED, MedDRA, RxNorm

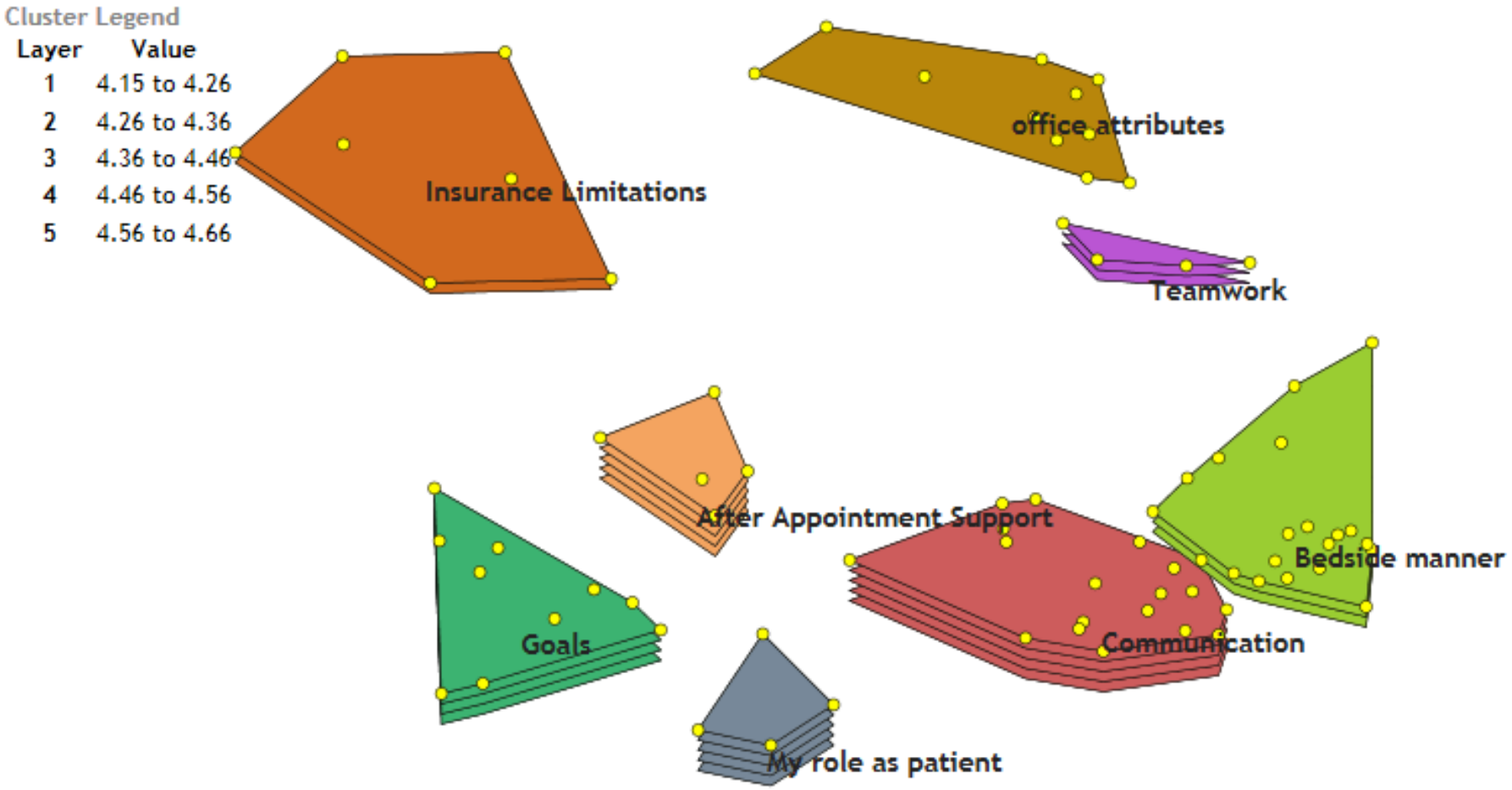
Conceptual Model of Good Health Care

- Concept mapping approach
- Survey PLM patients, vulnerable patients and stakeholders
- Target conditions: heart disease, cancer, stroke, diabetes, hypertension and arthritis
- Participation in phase one: concept elicitation
  - *187 PLM patients*
  - *17 stakeholders*
  - 1300 patient statements
  - *500 stakeholder statements*
  - *~ 250 unique keywords*

Example Statements

My doctor/provider makes eye contact with me
I am able to choose which provider(s) I want to see
I am able to contact my doctor's office with any needs, even between visits
The doctor/provider does not seem rushed
I am treated with respect
My doctor/provider is knowledgeable about my condition(s) and appropriate treatments for my condition(s)
The costs for office visits and treatments/medicine are reasonable
I get answers to all of my questions
My doctor/provider takes time to explain (diagnosis, treatment options, prognosis, side effects) in sufficient detail
My doctor/provider gets to know me
The doctor/provider is on time for appointments
Treatments are effective
My doctor/provider appreciates my input and asks my opinion

Cluster Map (8)



A New Approach
Patient Input into Measure Development

- Innovation in PRO-PM development process tested in collaboration with the NQF Measure Incubator™
  - Provide qualitative and quantitative patient experience data
  - Perspective on outcomes of greatest importance
  - Provide context and recommendations for measurement priorities from patients' perspective

- Measure Incubator PRO-PM projects
  - COPD - 2,545 PLM community
  - Multiple Sclerosis - 57,198 PLM community
  - Rheumatoid Arthritis - 9,950 PLM community

In-Depth Patient Experience Reports

Building on PLM's experience with FDA Patient-Focused Drug Development Initiative

- Composition of the patient community
  - *gender, age, race, ethnicity, diagnosis status, education level and insurance*
- Analysis of relevant patient data
  - *symptom severity, health related quality of life, comorbidities, medications, non-drug treatments*
- Survey results using measurement tools under consideration for performance assessment
  - *determine how these surveys match patient priorities and where gaps exist*
- Qualitative analyses of patient discussion forum posts
- Summary and recommendations

Report for NQF Measure Incubator Rheumatoid Arthritis Expert Panel

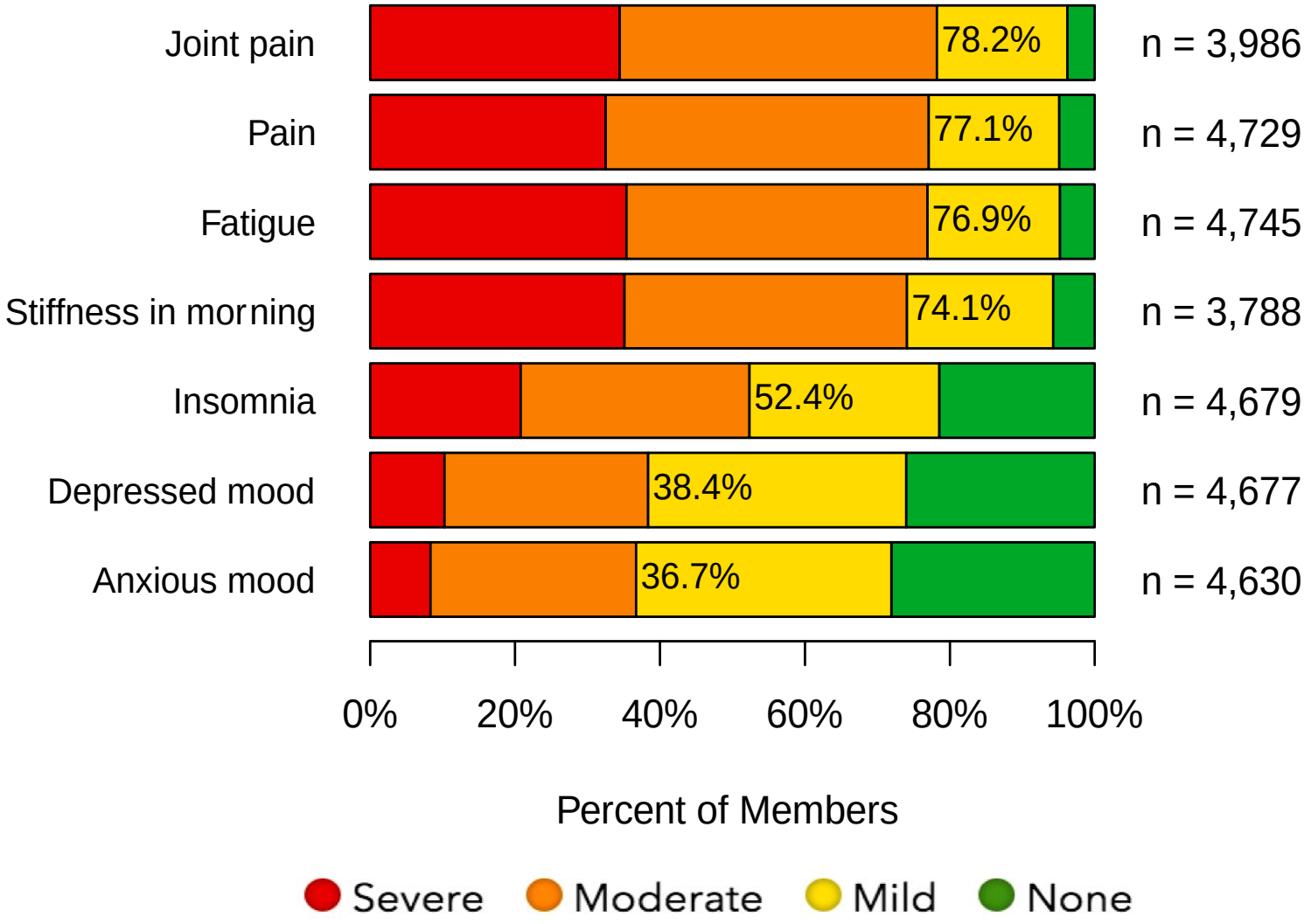
The Voice of the PatientsLikeMe Rheumatoid Arthritis Patient

A Brief Report on Patient Perceptions of Important Treatment Outcomes

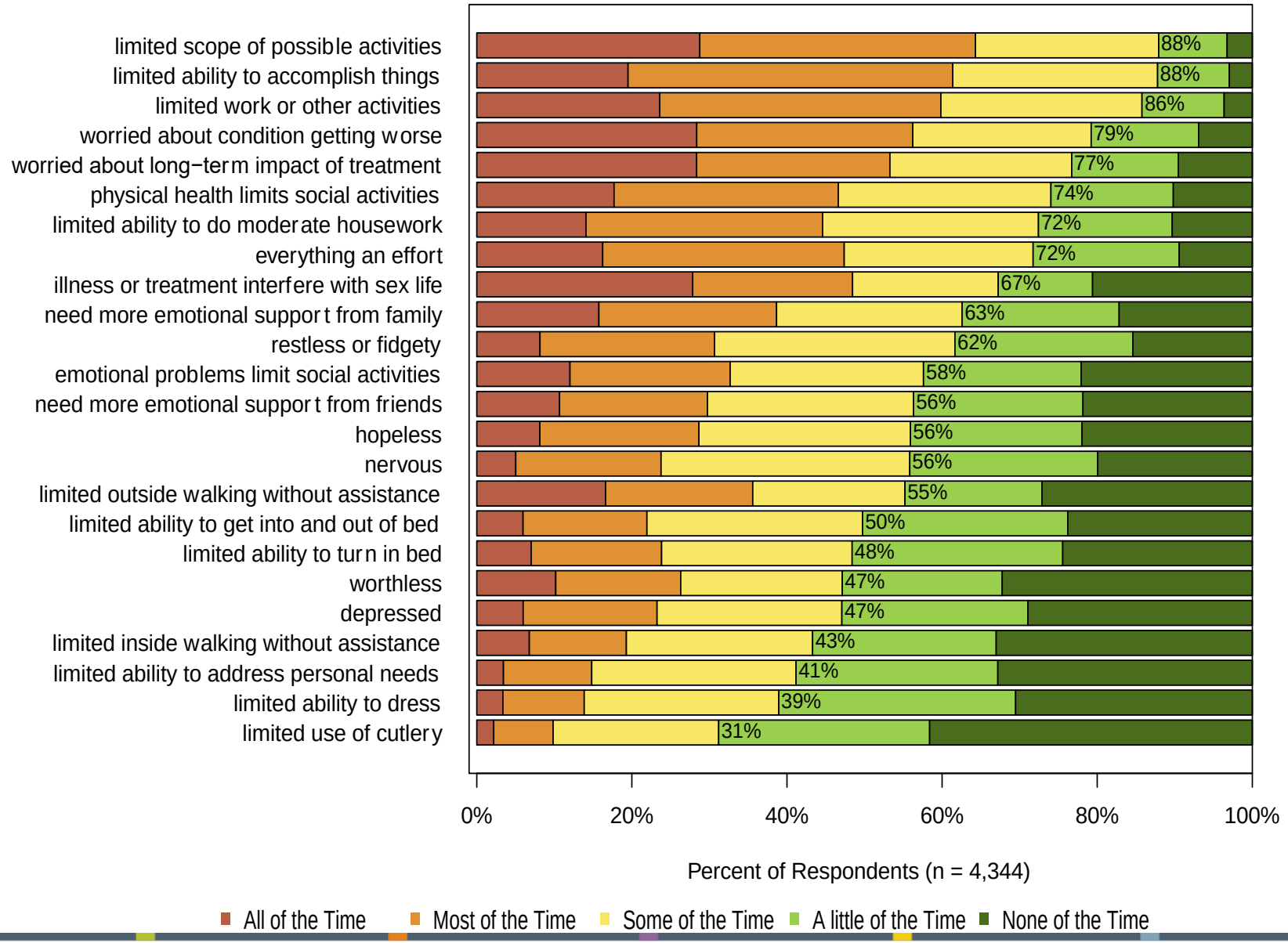
March 1st, 2017

patientslikeme®

Most Recently Reported Symptom Severity: RA Community



Most Recent Quality of Life Report: RA Community



RA Participant Survey (n=109)

- Clinical questions
  - *diagnosis, disease status*
- RAPID3 15 items
  - *When your doctor is trying to understand the impact that RA has on your life, how important is it that your doctor asked this question (1-5 scale)*
- RAPID3 content feedback
  - *Overall rating, content coverage, additional feedback*
- Demographic questions
  - *gender, age, race/ethnicity, and education*

RAPID3 Items Ranking

Table 7. Descriptive Statistics for RAPID3 Item Ratings

Item	% Endorsing "Not"	% Endorsing "A Little"	% Endorsing "Somewhat"	% Endorsing "Very"	% Endorsing "Extremely"
1: Dress	2.8	10.1	19.3	33.0	34.9
2: Get in and out of bed	3.7	10.1	15.6	35.8	34.9
3: Lift cup	5.5	11.0	27.5	28.4	27.5
4: Walk outdoors	5.5	3.7	15.6	41.3	33.9
5: Wash/Dry entire body	4.6	10.1	22.9	28.4	33.9
6: Bend down	2.8	11.0	19.3	35.8	31.2
7: Turn faucets	6.4	13.8	24.8	28.4	26.6
8: Get in and out of vehicle	3.7	6.4	22.0	37.6	30.3
9: Walk two miles	5.5	6.4	16.5	35.8	35.8
10: Recreational activities	4.6	8.3	19.3	32.1	35.8
11: Good night's sleep	1.8	1.8	10.1	28.4	57.8
12: Deal with anxiety	2.8	10.1	20.2	20.2	46.8
13: Deal with feeling blue	2.8	9.2	15.6	29.4	43.1
14: Pain	0.0	1.8	9.2	22.0	67.0
15: General functioning	0.0	1.8	10.1	26.6	60.6

Qualitative Feedback: What Content is Missing

- Difficulty completing daily activities
- Social/Recreational Activities (hobbies, sexual activity, religious activities)
- Occupational/ homemaker/ caregiving
- Other health-related concerns (appetite, muscle spasm, fatigue, appearance)
- Emotional Health/wellbeing
- RA-symptoms
- Cognitive function (concentration, brain fog)
- Treatment related concerns (medications, side effects)
- Environmental impacts
- Comorbid conditions

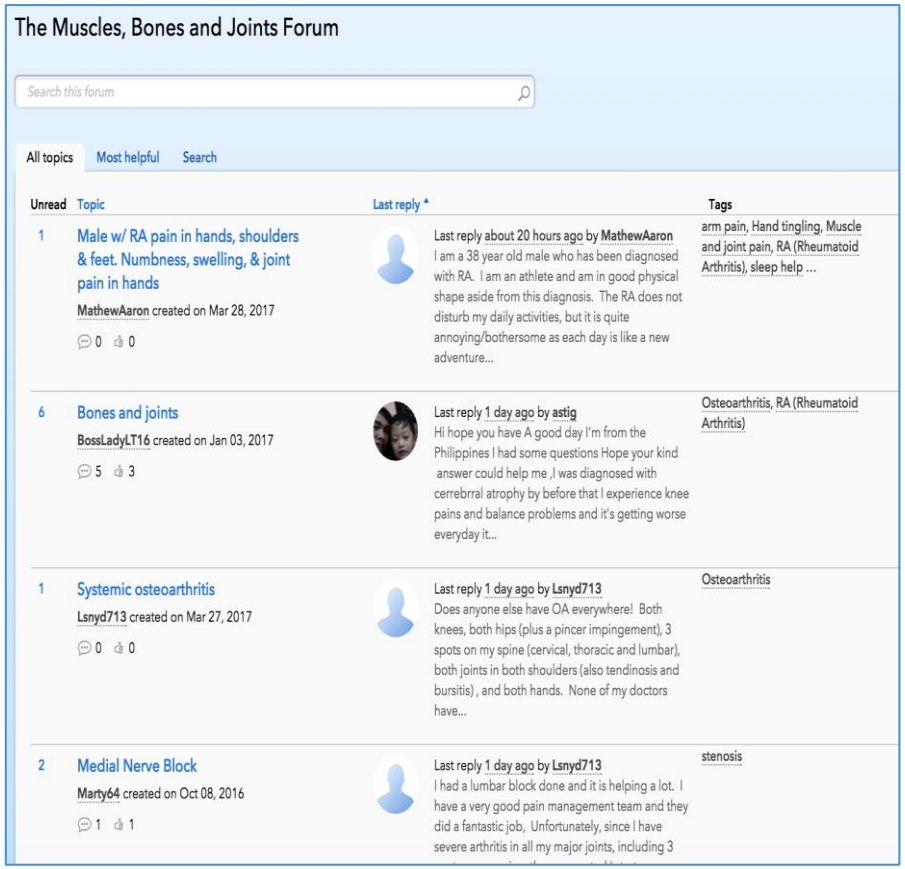
Qualitative Feedback: Additional

- Time period – one week
- Place to add information
- More targeted RA questions
- Responses that lead to referrals (depression)
- What happens to the information?
- How will it inform my treatment?

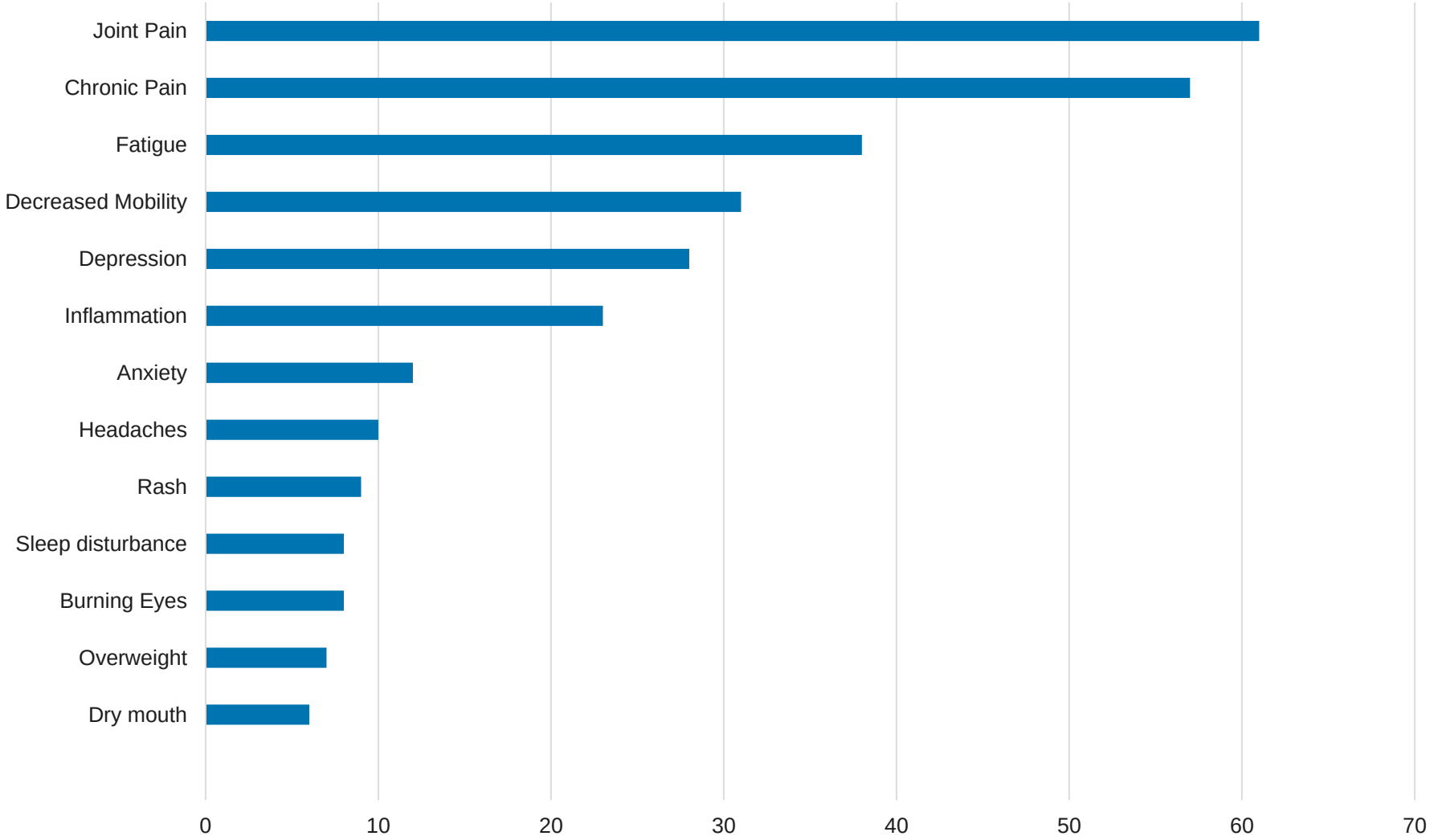
- Use as a semi-structured interview
- Vary questions depending on the severity
- Computer adaptive technology
- 2 participants created a tool
  - *20 items and 10 items*
  - *"Very Active Patient 1" or VAP1*

PLM RA Discussion Forum

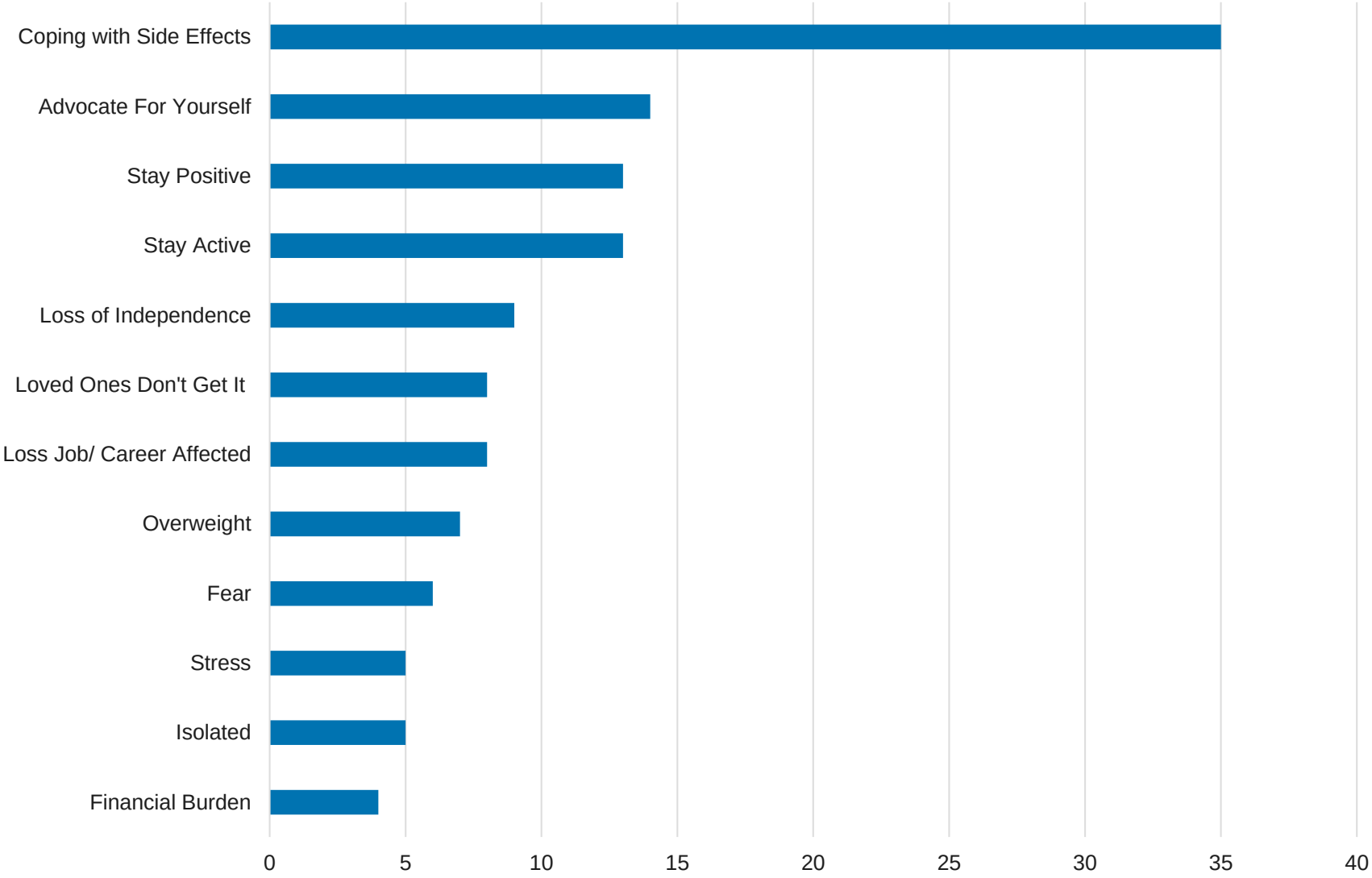
- Posts from Jan 2014 – Nov 2016
- 148 patients
- Average age 53.8 years
- 92% female
- 44% RA is primary condition
- Mention of outcomes/symptoms
- 377 posts analyzed and coded



Treatment Outcomes of Importance within PatientsLikeMe RA Community n=377



Topics of Importance Regarding Quality of Life Within the PatientsLikeMe RA Community n=377



What We are Learning

- Similarities exist across conditions in patient experience and priorities when considering health related quality of life (HRQL)
- Aggregate patient experience reports turn anecdotal input into quantifiable evidence
- Patients have very practical ideas about the value of measurement
- Relevant and easy to interpret data at point of care is necessary to support shared decision making
- Patient-centered context for measure developers is multifaceted and identifies:
  - What is meaningful to measure
  - How best to operationalize measurement
  - What are the most relevant uses of data by patients and clinicians
  - How to align with individual patient goals

Questions for MD Workshop

- Compared to current methods of eliciting patient input, how would this approach help you?
- What additional patient experience information would help inform measure development?
- At what point in the process of measure development would you expect this information to be most helpful?
- How do you think this method of patient input could change the overall assessment of quality and performance in health care?

Lunch

Person & Family Engagement in Quality Measurement

Rachel Johnson-DeRycke
Person & Family Engagement Team Lead
Yale Center for Outcomes Research & Evaluation (CORE)

NQF Measure Developer Workshop

May 17, 2017

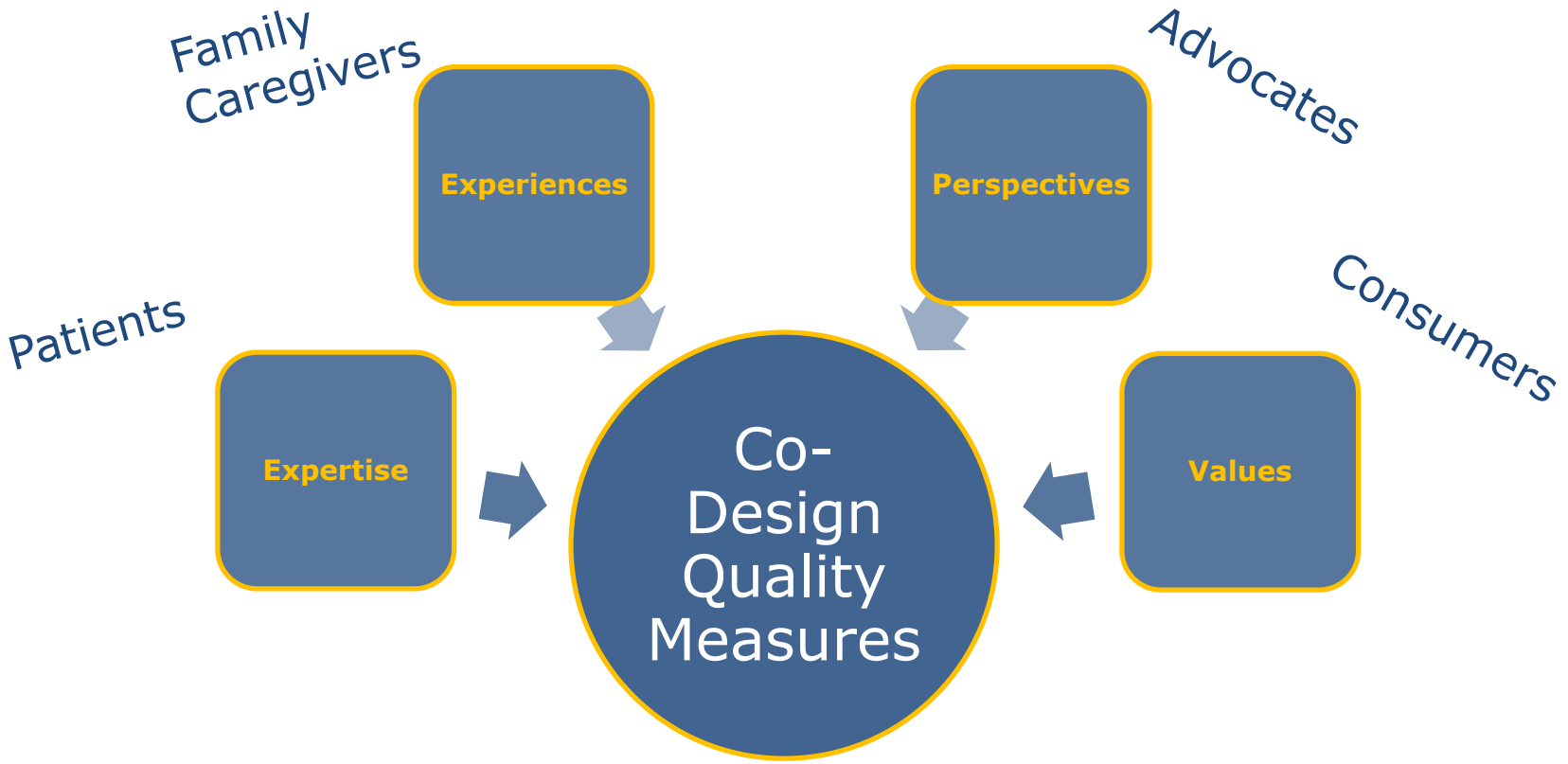
Key Players

national partnership for women & families

Definitions

PFE	Person and family engagement
Patients	Individuals managing a health condition who interact with the health care system
Family Caregivers	Individuals who assist a family member in managing their health and health care
Consumers	Individuals who have experience with the health care system but do not currently manage a health condition
Advocates	Individuals who work at nonprofit organizations representing a constituency of patients
PFE Partners	Patients, family caregivers, consumers, and advocates who collaborate with CORE on quality measurement projects
PFEN	Person & Family Engagement Network, made up entirely of PFE Partners

Persons & Families Can Shape Their Healthcare



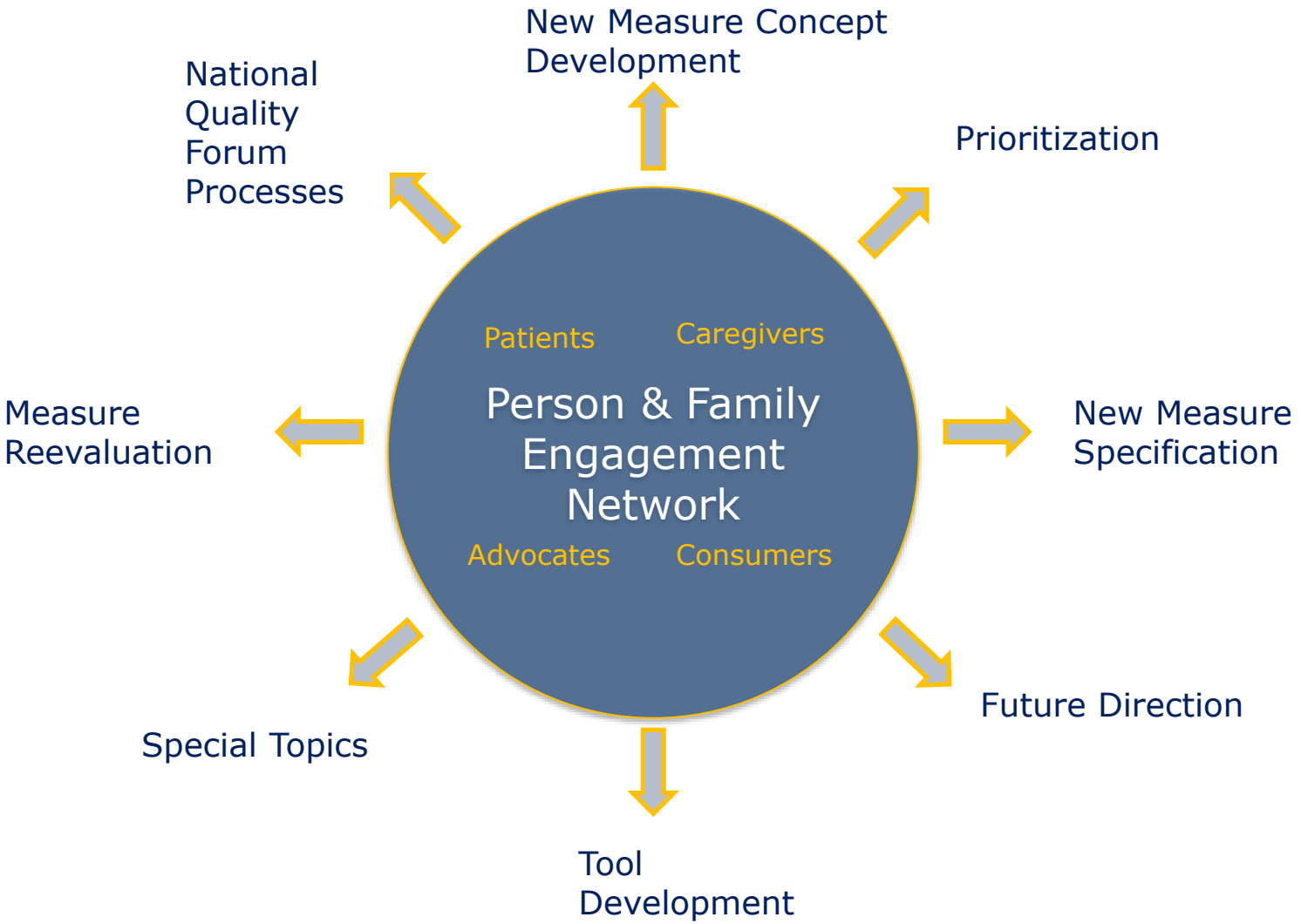
How PFE Partners Contribute

- Guide decision-making
- Recommend additional research or analysis
- Prioritize direction of measure development
- Define measure outcomes
- Create and refine tools
- Ensure measures are described and displayed so patients can understand them
- Teach us how to engage them effectively

Ways We Engage

- Technical Expert Panels
- Co-Development Groups
- Concept Advisory Groups
- Communication Workshops
- Surveys
- Interviews
- Discussion Forums and Webinars

Person & Family Engagement Network



PFE Impact

- Co-developed six new outcome measures
- Co-created Hospital Star Ratings methodology
- Co-developed preliminary conceptual relationship between socio-demographic factors and outcomes
- Refined, simplified, and improved patient-centeredness of *Hospital Compare* language and displays

Engagement Process

Challenges and Mitigation Strategies

Recruitment

Engagement

Organizational

What PFE Partners are Saying

- *"I definitely feel I have made an impact. [As a patient], I felt truly listened to for the first time."*
- *"It's great to have been part of something that will ultimately change the lives of patients and their families."*
- *"This [working group] was a unique experience in my experience as an advocate, and I hope this becomes standard operating practice in the measurement process."*
- *"It was valuable to have the dynamic discussion with the working group, and see the discussion progress…It was a good process. [I felt] we moved it forward in a way that's meaningful and improved the site, made it more useable."*
- *"[I am] excited about the project, it could help ensure that consumers are consistently engaged in their health decisions."*

What CORE Measure Developers are Saying

- *"Working with the patients and patient advocates has been invaluable …The feedback we have received has directly resulted in substantial changes to the methodology."*
- *"Given the diverse and uniquely relevant experiences of the patient advocates, their contributions had broad application for improving how we communicate about…outcome measures."*
- *"We greatly benefited from having…patient advocates on our technical expert panel. They brought a perspective and expertise that incorporated a truly patient-centered voice into early measure development and complemented our other content experts."*
- *"[Patients and advocates] brought different perspectives, which was critically important, and ensured the tool included dimensions that mattered to patients."*

Moving Forward: PFE and NQF

- Increase patient representation at NQF
  - PFE Committee
  - Patients on the NQF board
- Involve patients in preparing and presenting measures for endorsement
  - NQF measure submission forms
  - Co-present measures
- Encourage patient, family, advocate, and consumer involvement in endorsement processes
  - Proactive outreach about upcoming NQF meetings
  - Participation in public comment

Questions and Discussion

Hybrid Measures and the Core Clinical Data Elements (CCDE)

Agenda

- Background on CORE's Measures
- CORE's Approach to Measure Development
- Measure Specification and Testing
- Future Considerations

Why Use EHR Data in Measures

- Clinical data that are available and easily extracted for most patients can have broad utility in outcome measurement
  - Address provider's preference for clinical data in patient-level risk adjustment
  - Simplify development of risk-adjustment models and reduce overlap of work across developers
  - Align with CMS goals to use more clinical data in quality measures and programs

Measures Developed

- We have developed 3 measures of hospital performance on patient outcomes that use EHR data
  - Hybrid AMI 30-day mortality measure (#2473)
  - Hybrid Hospital-Wide Readmission measure (#2879)
  - Hybrid 30-day stroke mortality measure (#2877)

Hybrid Measures

- Use a combination of claims and EHR data
  - Claims to define the condition cohort usually for a specific set of conditions or procedures
  - Claims or other administrative data to define the outcome
  - EHR data in risk-adjustment

Structure of Hybrid Measures

	Cohort	Risk Adjustment	Outcome		New Risk Variables
Data Source	Claims	Claims	Claims or Enrollment Data		EHR
Data Type	ICD codes (grouped into condition categories)	ICD codes (grouped into condition categories)	ICD codes (grouped into condition categories) Disenrollment from Medicare	+	Clinical data captured during the episode of care
Data Elements	Principal discharge diagnosis	Principal & secondary diagnoses (12 months prior to and including the index admission)	Principal diagnoses and procedure codes Disenrollment due to death		First-Captured vital signs and lab test results

CENTER FOR OUTCOMES RESEARCH AND EVALUATION CORE

CMS CENTERS FOR MEDICARE & MEDICAID SERVICES

Principles for Selecting Data

- Begin by using EHR data in risk adjustment
  - Responsive to stakeholder concerns that claims data are not ideal
  - Data needed for outcomes are less feasible
- Data elements must be feasibly obtainable now
- Data elements that require changes to clinical workflow should be limited to those of critical importance and should come in later stages

Feasibility Criteria

- To be feasible, CORE determined that data elements must be:
  - Consistently obtained in the target population based on current clinical practice
  - Captured with a standard definition and recorded in a standard format
  - Entered in structured fields that are feasibly retrieved from current EHR systems

NQF Criteria

- Developers must provide a feasibility score for all data elements with supporting evidence
  - Data availability
  - Data accuracy
  - The presence of a standard coding system
  - Data availability in routine workflow
- The feasibility of measure logic must also be assessed

Core Clinical Data Elements

- Each of the hybrid measures uses a subset of 22 clinical data elements we have identified as feasible and predictive of mortality and readmission.
  - Vital signs
  - Laboratory test results
  - Must be the first value captured during the episode to reflect patients' status at the time of arrival

Core Clinical Data Elements

Data Elements	Units of Measurement	Time Window for First Captured Values
Patient Characteristics		
Age at admission	Years	---
Gender	Male or female	---
First-Captured Vital Signs		
Heart Rate	Beats per minute	0-2 hours
Systolic Blood Pressure	mmHg	0-2 hours
Diastolic Blood Pressure	mmHg	0-2 hours
Respiratory Rate	Breath per minute	0-2 hours
Temperature	Degrees Fahrenheit	0-2 hours
Oxygen Saturation	Percent	0-2 hours
Weight	Pounds	0-24 hours
First-Captured Laboratory Results		
Hemoglobin	g/dL	0-24 hours
Hematocrit	% red blood cells	0-24 hours
Platelet	Count	0-24 hours
WBC Count	Cells/mL	0-24 hours
Potassium	mEq/L	0-24 hours
Sodium	mEq/L	0-24 hours
Chloride	mEq/L	0-24 hours
Bicarbonate	mmol/L	0-24 hours
BUN	mg/dL	0-24 hours
Creatinine	mg/dL	0-24 hours
Glucose	mg/dL	0-24 hours
Troponin	ng/mL	0-24 hours

Other Measures

- We are also developing
  - Measures that use only data from EHRs
    - For cohort, risk adjustment, and outcome assessment
  - Patient-reported outcome measures
    - Will likely leverage EHRs for data collection

Approach to Measure Development

- We begin with rigorous assessment of the measure concept and feasibility
  - Technical experts
  - Person and Family Engagement
  - Review the literature
  - Previous feasibility assessments

Sources of EHR Data

- Development requires EHR data or a surrogate
- We have previously partnered with
  - Health IT vendors
  - Health Systems
  - Registries populated with chart abstracted data
- For hybrid measures we also require patient identifiers to link EHR data with claims

Development Steps

- Initial assessment of the feasibility of measure specifications including EHR data elements
- EHR data validity testing
- Electronic specification and testing measure logic

Initial Assessment of Feasibility

- Through direct testing of EHR data we establish that the data are
  - Structured (numerical or pseudo-numerical lists)
  - Captured on most patients in the target measure cohort
  - Function as expected in the measure calculation

Initial Data Validity Testing

- To insure that the data we use in testing is accurate
  - Work to understand what validity checks were done by the entity with the data, or
  - Perform validity checks in real time as we build an extracted dataset
- We attempt to do this in more than 1 EHR environment

Electronic Specifications

- Measure Authoring Tool (MAT) output
  - Human and machine readable logic expressed using accepted quality measurement standards
- Value sets for included data elements
  - Must be up to date and available through the Value Set Authority Center
  - Some proprietary coding systems cannot be downloaded for submission

MAT Testing

- Recruit volunteer hospitals
- Train Health IT staff on the specification and assist in developing a data query
- Review output and code
- Train abstractors to identify the same data elements from the clinical record
- Assess match between extracted (by automated query) and abstracted values
  - Redundant with data element validity testing

Considerations

- Is data element validity testing sufficient or must developers always test the measure logic (MAT output) as a separate step?
- What should be the standards for handling missing data in measures of performance used for accountability?
- What should the standards be for testing measure score reliability when most measures are developed and initially tested among a small number of providers?

May 17, 2017

Use of Electronic Health Record Data in Measure Development

National Quality Forum Measure Developers Workshop

Colleen M. McKiernan, MSPH CPH

Senior Consultant, Federal Health and Human Services

The Lewin Group

Madison Davidson, MPH CPH

Research Consultant, Federal Health and Human Services

The Lewin Group

Electronic Data are Transforming Measure Development

The conventional measure development process introduces data in later stages of the lifecycle

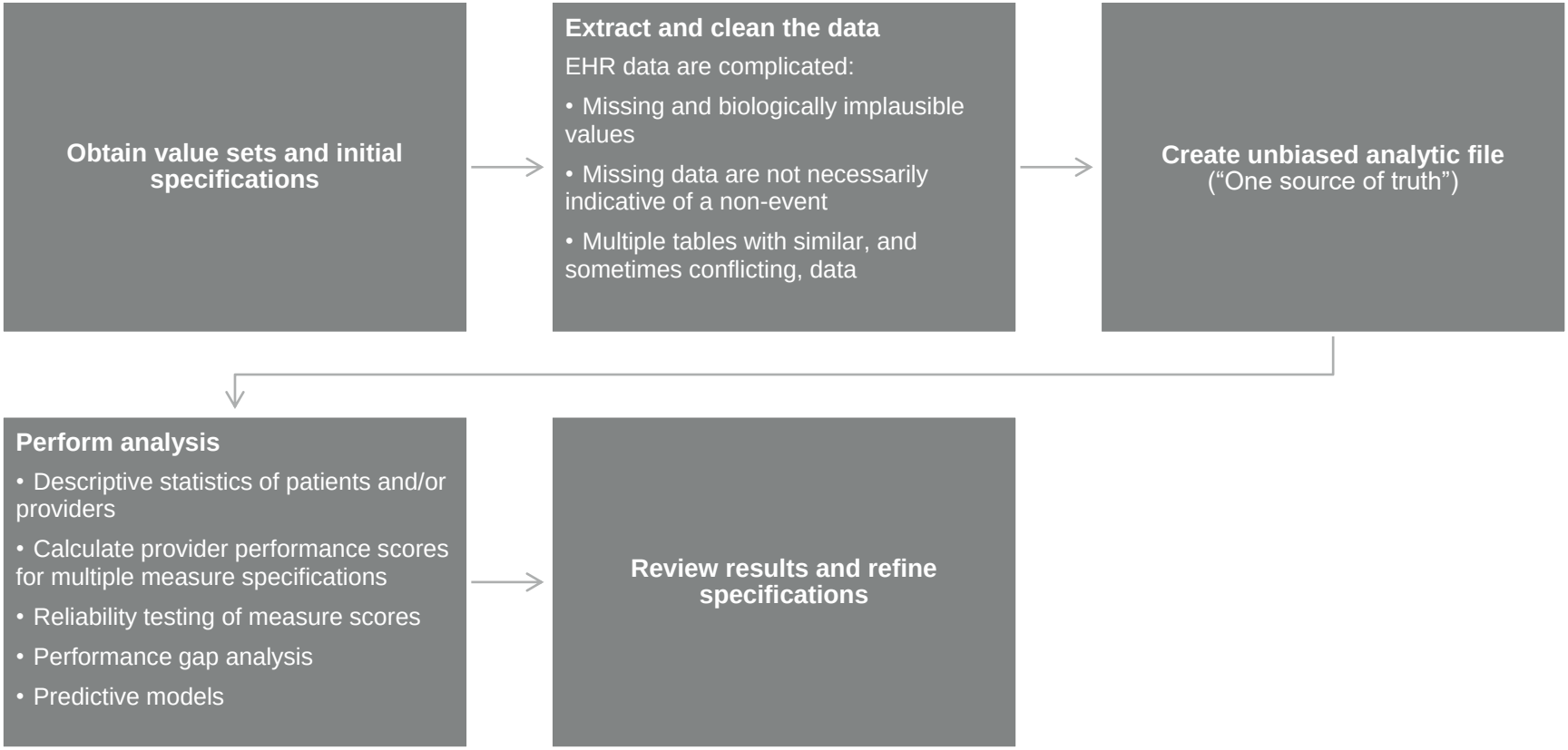
Easy access to large electronic datasets allows us to integrate data at multiple stages in during development, transforming the lifecycle into a data-driven process

Data are used during this stage

Use of Electronic Data has Benefits and Challenges

Benefits	Challenges
Big data: Samples are large, even for rare diagnoses *Cost-effective:* Reduces need for additional data contributors (including recruitment, contracting, and abstraction costs) *Easier concept evaluation:* Transforms measure development into a data-driven, 'fast-fail' approach *Fast:* Analyses take weeks, not months *Similar:* Data are similar to sources that will be used once a measure is implemented	*Convenience sampling:* Data are an opportunity sample not designed for quality measurement *Bias for contributors:* Natural bias is driven by the contributing EHRs, which can impact applicability of results (sociodemographic strata) *Complex:* Analytics may be deceptively complex, making it easy to mistake precision for accuracy or validity *Large samples could make the insignificant significant:* Because IPPs are so big, everything is statistically significant; these results may not be clinically meaningful

Using EHR Data for Measure Development and Testing



Draft Measure Specifications: HbA1C Targets in the Frail Elderly

- Due to risk of overtreatment leading to clinically significant hypoglycemic events, providers should take care when establishing HbA1c targets for management of frail or elderly patients with Type II diabetes (*Choosing Wisely*® | *ABIM Foundation*)
  - **The Challenge:** How do we collect sufficient data to explore different combinations of characteristics to identify patients appropriate for inclusion in this measure? What HbA1c score(s) signify increased risk of hypoglycemia?
  - **The Solution:** Using OptumOne data to obtain clinical history, HbA1c values, and other clinical indicators, we are able to provide a quick, cost-effective alternative to performing site testing of draft specifications

Draft Measure Specifications: HbA1C Targets in the Frail Elderly

OptumOne EHR Data Provided a Well-Powered Sample for Alpha Testing

	Type II Diabetics	Type II Diabetics with Pre-Defined List of Comorbidities
Sample Patient Population	139,627	102,776
Count of Unique Providers	18,679	17,798
Count of Unique Providers with at Least 20 Patients	2,868	2,266

- Use of these data allowed us to test ***128 different combinations*** of clinical characteristics and comorbidities to help define inclusion/exclusion criteria
- Access to these data also allowed us to test multiple numerator thresholds to refine measure specification

QUESTIONS?

Colleen McKiernan | Colleen.McKiernan@Lewin.com

Madison Davidson | Madison.Davidson@Lewin.com

NATIONAL QUALITY FORUM

Trial Use Program for eMeasures

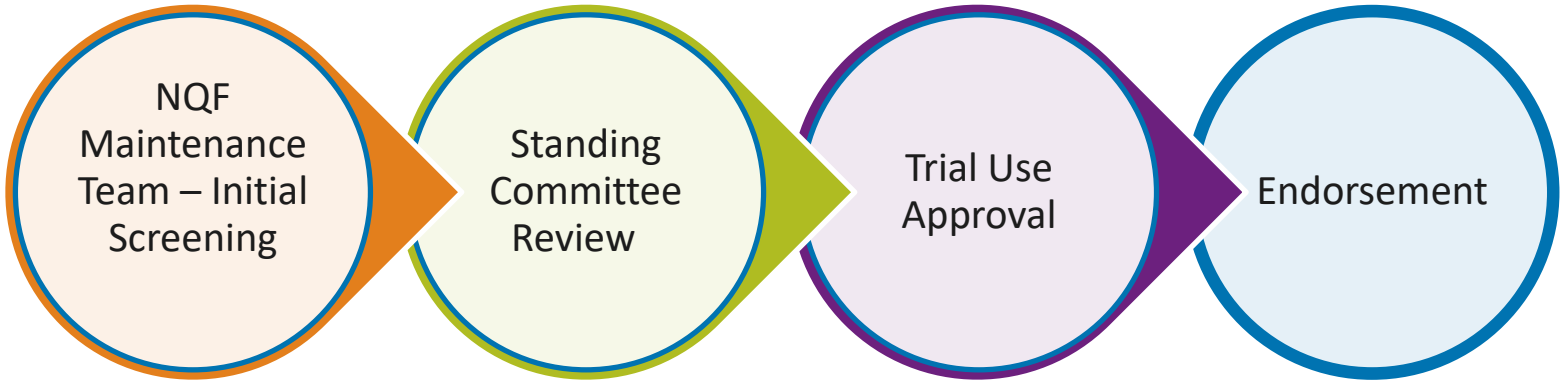
Kyle Cobb, Senior Director, NQF

May 17, 2017

The Need

- Interest in developing more eMeasures for Federal programs and obtaining NQF endorsement
- eMeasure may need to be more widely implemented to meet NQF endorsement criteria
- Trial Use Program is specifically for these type eMeasures ready for implementation, but cannot yet be adequately tested.

Trial Use Approval -> Endorsement



Screening Criteria

- ✓ Meet all criteria under *Importance to Measure and Report*
- ✓ eMeasure Feasibility Scorecard
- ✓ eMeasure Testing Form for Trial Use with
  - *results from BONNIE Tool simulated data set, or*
  - *test data set from another source*
- ✓ Plan for future use and discussion of how the measures will be useful for accountability and improvement
- ✓ Identification of related and competing measures and a plan for harmonization or justification for developing a competing measure

Trial Use Approval

NQF Criteria for Evaluation	eMeasure Testing Challenges	Trial Use Approval Evaluation Difference
Importance to Measure and Report		
Scientific Acceptability of Measure Properties	✓ Testing in 3 EHRs	✓ Reliability 2a1 only ✓ Validity 2b1 only
Feasibility	✓ Implementation readiness	
Usability and Use		
Related and Competing Measures		

Endorsement Evaluation Options

- Approved for Trial Use eMeasures may be submitted for endorsement at any time prior to the three-year expiration.
  - ***Option 1:*** *Submit and evaluate only the Scientific Acceptability of Measure Properties criterion, including the final eMeasure specification and all testing. If endorsed, endorsement date will assume the Approval for Trial Use date.*
  - ***Option 2****: Submit and evaluate on all criteria. If endorsed, a new endorsement date will be identified and endorsement maintenance will be scheduled from the new endorsement date, at which time it will be submitted for endorsement maintenance and subject to evaluation on all criteria.*

Break

Scientific Acceptability: Expectations and Examples

Karen Johnson, Senior Director, NQF

May 17, 2017

Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of health care delivery

2a. Reliability (must-pass)

2a1. Precise specifications

2a2. Reliability testing—data elements or measure score

2b. **Validity (must-pass)**

2b1. Specifications consistent with evidence

2b2. Validity testing—data elements or measure score

2b3. Justification of exclusions—relates to evidence

2b4. Risk adjustment—typically for outcome/cost/resource use

2b5. Identification of differences in performance

2b6. Comparability of data sources/methods

2b7. Missing data

Questions that a consideration of scientific acceptability helps us to answer

- Are the specifications clear so that everyone will calculate the measure in the same way?
- Are the specifications consistent with the evidence?
- Is the variation between providers primarily due to real differences? Or is it because there is a lot of "noise" in the measurement?
- Is the measure actually measuring what it is intended to measure (i.e., quality of care)?
- Do the results of the measurement allow for correct conclusions about quality of care?

Why is this important? Consider potential consequences of…

- Inconsistent measurement
- Inaccurate measurement
- Measurement that cannot differentiate between providers
- Measurement that leads to wrong conclusions about quality of care

Considerations

- Reliability and validity are not all-or-none properties
  - *They are a matter of degree*
- Reliability and validity are not static
  - *They vary with different conditions of using the measure*

- NQF allows flexible testing options (not prescriptive)
  - *Level of testing, data used in testing, methods of testing*
- NQF does not set specific thresholds for R/V
  - *Results should be within acceptable norms*
- Testing can be done on samples
- Prior evidence may be used as appropriate

Reliability

NQF Reliability Criterion – *Goal is to have consistent and precise measurement*

Criterion #2: Scientific Acceptability of Measure Properties

2a. Reliability (must-pass)

2a1. Precise specifications

2a2. Reliability testing—data elements or measure score

2b. Validity (must-pass)

Consistent (repeatable/reproducible) measurement

- Requires precise, unambiguous, and complete **specifications**
  - *Requirements vary based on type of measure*
- Can be demonstrated through empirical reliability testing of the **data elements**

Testing Data Elements for Reliability

- Uses patient-level data
- Needed for all critical data elements (not just computed score)
  - *At a minimum, numerator, denominator, exclusions*
- If based on instrument or scale (e.g., PRO-PM), reliability of the instrument or scale (e.g., internal consistency, test-retest analysis)
- If data element validity has been demonstrated, then additional demonstration of data element reliability is not required

Testing Data Elements for Reliability

- What we often get (for measures not based on instruments)
  - *Percent agreement (only)*
  - *Kappa statistics*
- What we'd like to see
  - *Kappa statistics or ICCs*
  - *Others??*

Testing Data Elements for Reliability

- What we sometimes get
  - *Very little explanation of the method*
  - *One value, with no explanation*
  - *No interpretation*
- What we'd like to see
  - *Detailed explanation of method*
  - *Values for all critical data elements—and if missing, why*
  - *Interpretation (e.g., use of Landis and Koch classification)*
    - » If values not great, speculate on why not; re-testing is expected

Not-So-Recent Example: Improvement in Ambulation -- Sample

- **Dataset**: Medicare Home Health Outcome and Assessment Information Set (OASIS), March-June 2009
- **Entities**: 20 home health agencies representing various types, locations, and sizes were included in the testing.
  - *Agencies were recruited through national and state associations and were selected based on capacity to do the reliability study*
  - *Ownership: 4 private, for-profit; 2 public for-profit chain; 6 private nonprofit; 1 health dept.; 5 hospital-based; 2 visiting nurse associations*
  - *Location: 4 states (AZ, MO, NY, TX)*
  - *Size: Less than 10,000 visits/year (n=3); 10,000-30,000 (n=10); greater than 30,000 (n=7)*
- **Patients**: 20-40 patients per agency, for a total of 500 patients
  - *Patient case-mix characteristics in the testing sample were similar to national averages with few significant differences*

Not-So-Recent Example: Improvement in Ambulation -- Methods

- Inter-rater reliability was assessed for the critical data elements used in this measure to determine the amount of agreement between 2 different nurses' assessments of the same patient.
- Patients were randomly selected from the planned visits for start or resumption of care and discharge assessments for each day of the study.
- The first nurse assessed the patient on the usual visit. The second nurse visited the patient and conducted the same assessment within 24 hours of the first assessment so as not to confound differences in assessment with real changes in the patient.
- Data analysis included:
  - *Percent agreement*
  - *Kappa statistic to adjust for chance agreement for categorical data*

Not-So-Recent Example: Improvement in Ambulation -- Results

Data Element	N	Percent Agreement	Kappa
Functional status score for ambulation	500	85%	0.62
Functional status score for ambulation prior to this start/resumption of care	495	83%	0.55
Primary diagnosis major diagnostic category	500	90%	0.70
Pain scale	500	88%	0.69
Location prior to this start/resumption of care	500	91%	0.72

Not-So-Recent Example: Improvement in Ambulation -- Interpretation

- A statistical measure of inter-rater reliability is Cohen's Kappa which ranges generally from 0.0 to 1.0 (although negative numbers are possible) where large numbers mean better reliability and values near zero suggest that agreement is attributable to chance alone
- Landis & Koch, 1977 offers the following classification of Kappa Interpretation
  - *< 0 Poor agreement*
  - *0.00 – 0.20 Slight agreement*
  - *0.21 – 0.40 Fair agreement*
  - *0.41 – 0.60 Moderate agreement*
  - *0.61 – 0.80 Substantial agreement*
  - *0.81 – 1.00 Almost perfect agreement*
- The results of our inter-rater analysis ranged from Kappa = 0.55 to 0.72. All values except one were in the range of substantial agreement and one was in the range of moderate agreement, demonstrating acceptable reliability of the assessment data used in the performance measure

Precise measurement

- Measurement that allows one to differentiate between providers
  - *Critical for accountability applications*
- Is demonstrated through empirical reliability testing of the **measure score** (the results of the measure calculation)
  - *Uses data that have been aggregated across providers*
  - *Must be tested for the measure as specified*
    - » For example, if specified at the clinician and hospital levels of analysis, must have testing for both

Testing Measure Scores for Reliability

- What we often get (for measures not based on instruments)
  - *Signal-to-Noise analyses*
    - » Based on beta-binomial model
  - *Split-sample correlations*
  - *Stability analysis (results over time)*
- What we'd like to see
  - *Signal-to-Noise analyses*
    - » Various methods (beta-binomial; ANOVA)
  - *Other??*

Testing Measure Scores for Reliability

- What we sometimes get
  - *Very little explanation of the method*
  - *Results that exclude low-volume providers*
  - *No interpretation*
- What we'd like to see
  - *Detailed explanation of method*
  - *Testing for measure as specified*
    - » If you are not excluding low-volume providers from the measure, do not exclude them from testing
  - *Reliability estimates for providers of different sizes*
  - *Interpretation*
    - » If values not great, speculate on why not; re-testing is expected

Validity

NQF Validity Criterion – *Goal is to make valid conclusions about quality*

Criterion #2: Scientific Acceptability of Measure Properties

2a. **Reliability (must-pass)**

2b. **Validity (must-pass)**

- *2b1. Specifications consistent with evidence*
- *2b2. Validity testing—data elements or measure score*
- *2b3. Justification of exclusions—relates to evidence*
- *2b4. Risk adjustment—typically for outcome/cost/resource use*
- *2b5. Identification of differences in performance*
- *2b6. Comparability of data sources/methods*
- *2b7. Missing data*

NQF Validity Testing Requirements

Source: 2016 Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement, pp 15-16

2b2. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

NOTE: Validity testing applies to both the data elements and computed measure score. Validity testing of **data elements** typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the **measure score** include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). **Face validity** of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

Validity Testing: Key Points

Validity refers to the correctness of measurement.

- Empirical analysis is required for the measure as specified
  - *Data element – assesses the correctness of the data elements compared to a "gold standard"*
  - *Measure score – assesses the correctness of conclusions about quality that can be made based on the measure scores (i.e., a higher score on a quality measure reflects higher quality).*
- Face validity of the measure score also is accepted

Testing Data Elements for Validity

- Uses patient-level data
- Needed for all critical data elements (not just computed score)
  - *At a minimum, numerator, denominator, exclusions*
  - *Should expect high level of agreement*
- If based on instrument or scale (e.g., PRO-PM), validity of the instrument or scale (e.g., content validity, confirmatory factor analysis)
- If data element validity has been demonstrated, then additional demonstration of data element reliability is not required

Testing Data Elements for Validity

- What we often get
  - *Percent agreement (only)*
  - *Kappa statistics*
- What we'd like to see
  - *Sensitivity*
  - *Specificity*
  - *Positive predictive value*
  - *Negative predictive value*

Testing Data Elements for Validity

- What we often get
  - *Very little explanation of the method*
  - *One value, with no explanation*
  - *No interpretation*
- What we'd like to see
  - *Detailed explanation of method, with gold standard plainly stated*
  - *Values for all critical data elements—and if missing, why*
  - *Interpretation (e.g., use of Landis and Koch classification)*
    - » If values not great, speculate on why not; re-testing is expected

Face Validity of the Measure Score

- Judgment of whether, on the face of it, the measure results appear to reflect quality of care
  - *Subjective determination*
  - *Weakest form of validity—therefore subject to challenge*
- NQF requirements
  - *Systematic assessment of the score from the measure as specified*
  - *Assesses whether measure results are an accurate reflection of performance on quality or resource use*
  - *Assesses whether measure results can distinguish good from poor performance*

Face Validity of the Measure Score

- What we sometimes get
  - *Assessments of whether the measure is a "good idea"*
  - *Feedback on the construction of the measure*
    - e.g., exclusions, ICD codes, etc.
  - *Feedback on the feasibility of the measure*
  - *No description of the experts involved*
  - *Little to no description of what was assessed*
  - *A general statement of results (e.g., the panel agreed the measure is valid)*
  - *No results at all*

Face Validity of the Measure Score

- What we'd like to see
  - *Details about what was asked and how the assessment was done*
  - *A list of the experts with their credentials*
    - » NQF doesn't require a particular number of participants
    - » We'd prefer that you query individuals NOT involved in the development process
  - *Actual results of the assessment*
    - » 9 of the 10 experts strongly agreed that the measure results differentiate good from poor quality of care; none disagreed with that statement

Empirical Testing of the Measure Score

Our thinking has evolved…

- ***THEN***: *Different types of validity (e.g., predictive, criterion, concurrent, convergent, construct)*
- ***NOW***: *Different ways to validate, but all involve assessing hypothesized relationship(s) of the measure results to results of another measure(s), based on knowledge of the underlying construct(s)*

- Very much a theoretical exercise
  - *Typically, not the role of a statistician (although they participate in identifying appropriate statistical testing and sample sizes and perform the mechanics of the testing)*

Empirical Testing of the Measure Score

- **Step 1**: Link concept of interest to some other concept by a "hypothesis" or construct (a "mini-theory")
  - *Usually there are many hypotheses than can be articulated*
  - *The hypothesis should indicate the direction of the relationship*
  - *Because the hypothesis is based on a theoretical framework, there should be some idea about the expected strength of the relationship*
- **Step 2**: Assess the "hypothesis" empirically
- **Step 3**: Examine the results of the testing
  - *Statistical significance, strength (effect size), direction*
  - *If the expected relationship is found, then it is likely that the hypothesis is sound and validity has been demonstrated to some extent*
  - *If the expected relationship is not found, then either hypothesis or measure (or both) is at fault*

Common Methods/Statistics

- Methods/statistics used in validation include:
  - *Correlations*
  - *Regression estimates (e.g., odds ratios)*
  - *Difference between means*
  - *Sensitivity/specificity; positive (and negative) predictive values (generally used for validity testing of data elements)*
- Choice of statistic depends on purpose (test difference, test relationship), testing level (data element, performance measure score), and type of data (continuous, categorical)

Recent Example: NICU Admission Temperature of Low-Birthweight Babies

- Hypothesis: Hospitals that have more babies with low temps at NICU admission have a higher NICU mortality rate
- Results:

	Proportion Cold/ Very Cool	Mortality
Hospital A	0.344	0.051
Hospital B	0.381	0.104
Hospital C	0.414	0.131

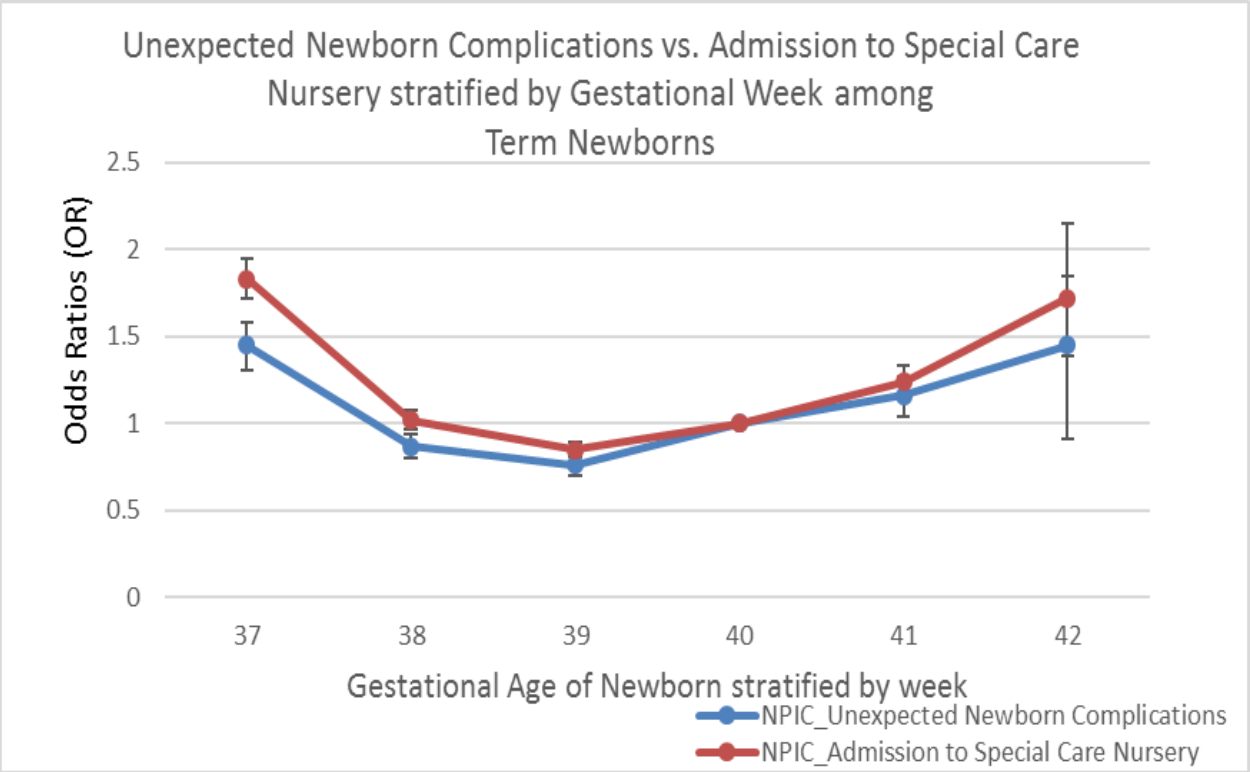
- Interpretation: Findings consistent with our expectations

Recent Example: Unexpected Complications in Term Newborns

- Hypothesis: We predict that our measure will be highly correlated with a neonatal admission to the NICU
- Results:

Gestational Week	Unexpected Newborn Complications*	Admission to Special Care Nursery (NICU)
37	1.45 (1.32, 1.59)	1.83 (1.72, 1.95)
38	0.87 (0.80, 0.94)	1.02 (0.97, 1.08)
39	0.76 (0.71, 0.82)	0.85 (0.81, 0.89)
40	1	1
41	1.16 (1.05, 1.28)	1.24 (1.15, 1.33)
42	1.45 (1.05, 1.99)	1.72 (1.39, 2.15)

Recent Example: Unexpected Complications in Term Newborns



Recent Example: Unexpected Complications in Term Newborns

- **Interpretation**: We can therefore conclude that severe and moderate morbidity in the NPIC cohort of newborns was validated by a very similar pattern of neonatal intensive care admissions with the lowest morbidity and NICU admissions being at 39 and 40 weeks and the highest at 37 and 42 weeks.
  - *This demonstrates that our metric successfully captures and quantifies neonatal morbidity in term newborns.*
- Our results are also comparable to published studies on the differences in neonatal morbidity by gestational week (3-4) with the lowest adverse neonatal outcomes occurring at 39 and 40 weeks of gestation compared with later weeks of term gestation.
- Additionally, in a separate test of correlation, both sets of odds ratios demonstrated high positive correlation further reinforcing the results and conclusions.

Empirical Testing of the Measure Score

- What we often get
  - *Naming of a method*
    - e.g., predictive, concurrent, convergent, construct validity
  - *A correlation table, but no explanation and little interpretation*
- What we'd like to see
  - Description of the hypothesized relationships
    - *Which other measures? Why? Expected direction? Expected strength?*
  - Interpretation of findings
    - *How does (or doesn't) this validate a measure?*

Considerations for Evaluation of Validity Testing

- Level of testing
  - *Patient-level data*
  - *Provider-level performance measure score*
- As with reliability, validity is not an not all-or-none property: it is a matter of degree
  - *Therefore, evaluation of validity requires judgment*
    - » Appropriate method
    - » Adequate sample (representative, numbers, patients & entities)
    - » Adequate results (within norms)
    - » Threats adequately assessed and accounted for
- Validity is not static: it can vary with different conditions of using the measure
  - *NQF requires a minimal demonstration of validity*
  - *Understanding that validity "in the field" may be less than what is seen in testing*

Extent of Validation

- Often, demonstration of validity should come from a series of studies
  - *More studies needed if no gold standard*
  - *More studies (with expected results) strengthens evidence of validity*
    - » As with clinical evidence, validity is built over time with multiple studies
  - *Even one study with unexpected results with respect to a gold standard may invalidate a measure*
- NQF criteria also require consideration of potential threats to validity (e.g., exclusions)

Threats to Validity

- Patients inappropriately excluded from measurement
- Differences in patient mix (for outcome and resource use measures)
- Measure scores that are generated with multiple data sources/methods
- Systematic error
  - *Systematically missing data*
  - *Systematically "incorrect" data*
  - *Both can be intentional or unintentional*

Threats to Validity: What We'd Like To See

- Exclusions
  - *At minimum, frequency of occurrence*
  - *Variability across measured entities*
  - *Preferably, sensitivity analysis of results with and without exclusions*
- Risk-adjustment (for outcome, resource use, some process measures)
  - *Empirical analysis to demonstrate that it isn't needed* ***OR***
  - *Conceptual and empirical approach*
    - » Conceptual rationale for SDS factors, even if SDS factors not ultimately included in the approach
  - *Discrimination and calibration statistics*

Threats to Validity: What We'd Like To See

- Meaningful differences
  - *At minimum, demonstrate variation among measured entities*
  - *Statistical analysis*
- Comparability if multiple data sources/methods are used
- Missing data
  - *How handled*
  - *How frequent*
  - *Demonstration that results are not biased*
    - » All particularly important for eMeasures, composite measures, or PRO-PMs

Questions for Discussion

- Any surprises?
- Score-level reliability – additional methods?
- Data element validity
  - *Do you typically validate the data elements?*
  - *Sensitivity/specificity/PPV/NPV not common. Why not?*
- Face validity
  - *Do you think NQF should drop face validity option?*
- Empirical testing of the measure score
  - *Does thinking of this as "assessing relationships" make this any clearer?*
  - *What are some of the barriers to score-level validation?*
  - *How can these be mitigated?*
- If you could change anything we do regarding validity, what would it be?

THANK YOU

NATIONAL QUALITY FORUM

Next Steps

Jean-Luc Tilly, Project Manager, NQF

May 17, 2017

Next Steps

- May 18-19: NQF Kaizen
  - More predictable and frequent submissions, faster review, easier access to information
  - Thank you for completing the Developer Survey!
  - Look for published follow-up report/public comment period
- Look for new project announcements on the NQF website, and from measuremaintenance@qualityforum.org
- Measure Developer Webinar June 17 (1:00 PM ET)