

## Welcome and Setting the Stage

Patti Brennan, University of Wisconsin-Madison  
Joyce Dubow, AARP



NATIONAL  
QUALITY FORUM

## Meeting Objectives

- Discuss the major methodological issues related to reliability and validity when aggregating PROM data into a performance measure;
- Identify unique considerations in relation to the NQF endorsement criteria for PRO-based performance measures (PRO-PM) (as compared to other quality outcome performance measures); and
- Lay out the critical path from PROM to PRO-PM endorsed by NQF for use in accountability and performance improvement.

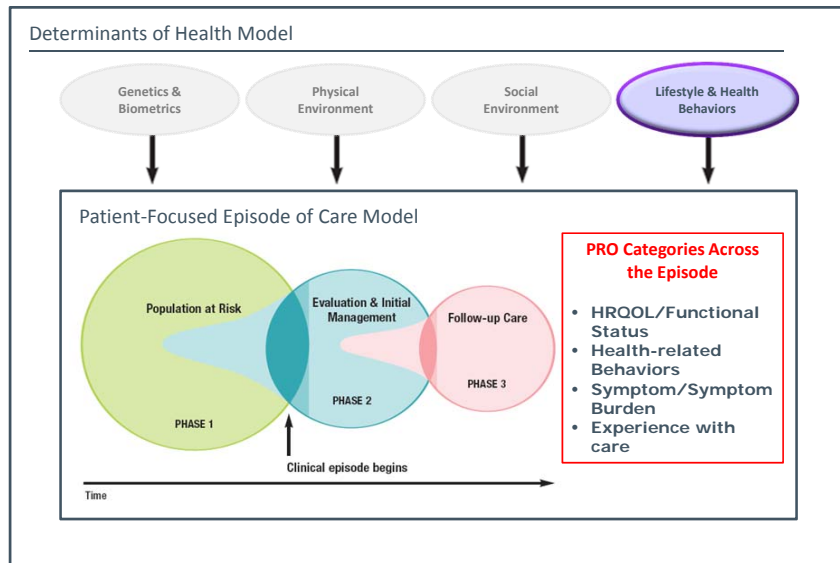
## Terminology

- **Patient-reported outcome (PRO):** The concept of any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.
- 
- **PRO measure (PROM):** Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).

## Terminology

- **Performance measure:** Numeric quantification of healthcare quality for a designated accountable healthcare entity, such as hospital, health plan, nursing home, clinician, etc.
- **PRO-based performance measure (PRO-PM):** A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

## Framing PROs Within Existing Conceptual Models



NATIONAL QUALITY FORUM

## NQF Endorses Performance Measures

- NQF endorses PRO-PMs, not the PROMs
- NQF endorses PRO-PMs for use in accountability applications such as public reporting and payment as well as improvement
- NQF evaluates suitability for endorsement based on a set of [evaluation criteria](#)
  - Importance to measure and report
  - Scientific acceptability of measure properties
  - Feasibility
  - Usability and use

NATIONAL QUALITY FORUM

6

## Four Major Endorsement Criteria Hierarchy and Rationale

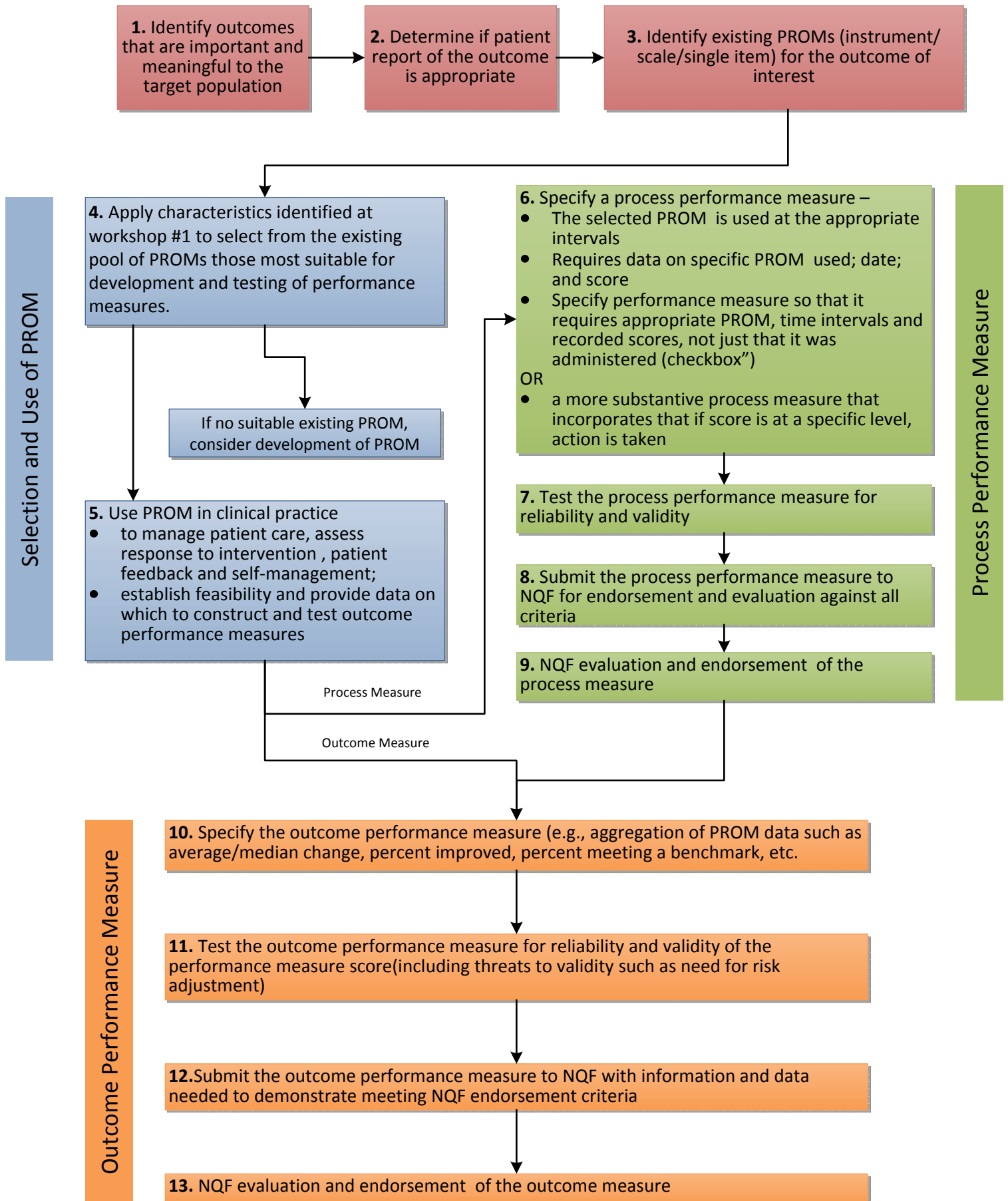
- **Importance to measure and report** – measure those aspects with greatest potential of driving improvements in healthcare and health (*evidence of effectiveness, performance gap, impact*) ; if not important, the other criteria less meaningful (*must-pass*)
- **Scientific acceptability of measure properties** – goal is to make valid conclusions about quality; if not *reliable and valid*, risk of improper interpretation (*must-pass*)
- **Feasibility** – ideally, cause as little burden as possible; if not feasible, consider alternative approaches
- **Usability & Use** – goal is to use endorsed performance measures for decisions related to both *accountability and improvement*

## Draft Pathway from PROM to NQF-endorsed PRO-PM

- Product of the workshop
- Draft pathway will be refined on Day 2
- [PRO Day1 am0900 Day2 am1050 Pathway.pdf](#)

## Timeline

- Call for nominations closed 4/2/12
- Hold workshop #1 7/30-31/12
- Expert Panel to discuss revision of first commissioned paper 8/21/12
- Receive final version of first commissioned paper and prepare draft report of findings/recommendations 8/31/12
- Hold workshop #2 9/11-12/12
- Expert Panel review 2<sup>nd</sup> paper revisions & draft report for comment 10/11/12
- Public/member comment period open 10/23 11/21/12
- Expert Panel to review comments received 12/3/12
- CSAC and NQF Board review and approval 12/20/12



PATHWAY NOTES – Correspond to Pathway Elements on Page 1

**NQF Criteria**  
Importance to Measure and Report

If patient/person is not the best source of information for the outcome, then explore clinical data and measurement

PRO refers to the concept  
PROM refers to the instrument, scale, or single-item to measure the PRO concept  
PRO-PM refers to PRO-based performance measure  
Many PROMs developed and tested (reliability, validity, responsiveness, identification of meaningful differences, etc.) primarily for research

**NQF Criteria**  
Importance to Measure and Report, 1c.  
Evidence - responsive to clinical intervention

Scientific Acceptability of Measure Properties:  
2a. Reliability and 2b. Validity  
Reliability and validity of data elements used in performance measure (i.e., PROM data)

**Characteristics for Selecting PROMs Identified in Commissioned Paper (Table 3)**  
1. Conceptual and Measurement Model Documented  
2. Reliability  
2a. *Internal consistency (multi-item scales)*  
2b. *Reproducibility (stability over time)*  
3. Validity  
3a. *Content Validity*  
3b. *Construct and Criterion-related Validity*  
3c. *Responsiveness*  
4. Interpretability of Scores 8  
5. Burden  
6. Alternatives modes and methods of administration  
7. Cultural and language adaptations  
8. Electronic health record (EHR) capability

Additional Characteristics from workshop:  
Meaningful · Actionable · Able to facilitate shared decision-making · Implementable

Process measure considered an interim step to encourage use and obtain data and experience so that outcome performance measure could be developed, tested, and endorsed

**NQF Criteria**  
2a.1 Precise specification  
2b.1 Specifications consistent with evidence

Should be specified so that data can be used to construct and test future outcome performance measures

Should be specified so that it is more than a “checkbox” – “checkbox” measures generally do not pass Importance to Measure and Report because not proximal to desired outcomes; doing an assessment is first step but far from sufficient to influence outcomes

**NQF Criteria**  
2a2 Reliability testing  
2b2 Validity testing

Is using a reliable and valid PROM sufficient demonstration of reliability and validity at the data element level? Is testing at the level of the performance measure needed?

**NQF Criteria**  
Importance to Measure and Report  
Scientific Acceptability of Measure Properties  
Usability and Use  
Feasibility

Does endorsement of performance measure increase use and provide more data/experience to develop and test outcome performance measure?  
Or does it divert focus and resources from outcome?

**NQF Criteria**  
Usability and Use  
Feasibility

**NQF Criteria**  
2a.1 Precise specification  
2b.1 Specifications consistent with evidence

**NQF Criteria**  
2a2. Reliability testing of performance measure score, e.g., signal-to-noise analysis (or is data element sufficient?)  
2b2. Validity testing of the performance measure score i.e., can make correct conclusions about quality of care (or is data element sufficient?)  
2b3. Exclusions justified  
2b4. Differences in case-mix (is risk adjustment needed, adequate?)  
2b5. Performance measure score discriminates among the accountable entities  
2b6. Comparability of different data sources/methods

**NQF Criteria**  
Importance to Measure and Report  
Scientific Acceptability of Measure Properties  
Feasibility  
Usability and Use

Lessons from the Field –  
Using PRO-PMs for  
Accountability



NATIONAL  
QUALITY FORUM

1

Lessons from the field – using PRO-PMs for  
accountability (public reporting, payment)

**Moderator:** Greg Pawlson

**Panel:**

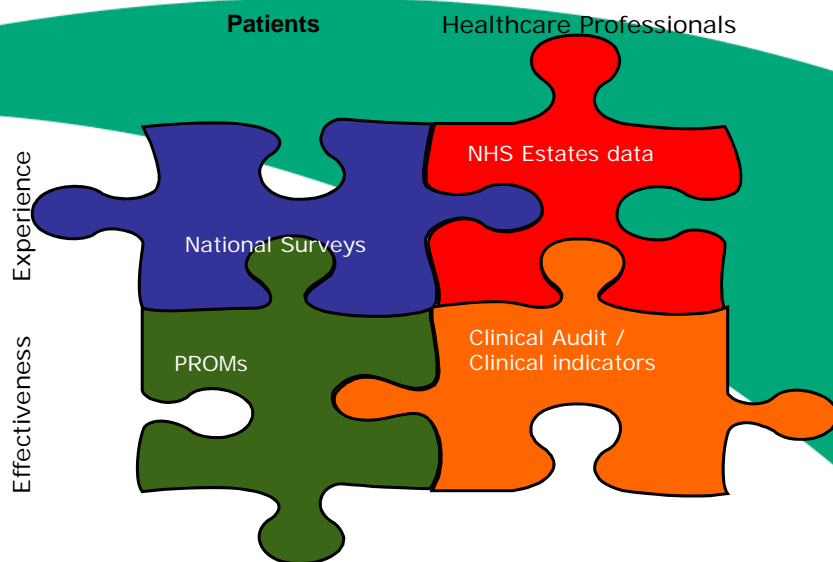
- **England**–**David Nuttall**, *Branch Head - Choice, AQP & PROMs, Strategy, Finance and NHS Directorate, Department of Health;*
- **Medicare Advantage**–**Elizabeth Goldstein**, *Director Division of Consumer Assessment and Plan Performance, Centers for Medicare & Medicaid Services*
- **Sweden**– **Stefan Larsson**, *Senior Partner & Managing Director Stockholm Office, Boston Consulting*



## Lessons from the field: Patient Reported Outcome Measures (PROMs) in the National Health Service

David Nuttall  
Department of Health, London  
National Quality Forum workshop  
11<sup>th</sup> September 2012

The purpose of collecting PROMs is to **Complete** the quality “picture”



## Introduction of PROMs into the NHS followed extensive piloting



- Collection of Patient Reported Outcome Measures (PROMs) was piloted from 2005 in a number of providers of NHS funded services. The results were published in 2007:
  - [http://www.rcseng.ac.uk/surgical\\_research/ceu/docs/POIS\\_Report\\_finalOct2011.pdf](http://www.rcseng.ac.uk/surgical_research/ceu/docs/POIS_Report_finalOct2011.pdf)
- Following on from the successful pilot, routine collection of PROMs data for NHS funded care was made mandatory and began for four procedures from April 2009:
  - Hip Replacement
  - Knee Replacement
  - Varicose Vein Surgery
  - Groin Hernia repairMore info:  
[http://www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/@dh/@ent/documents/digitalasset/dh\\_092625.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ent/documents/digitalasset/dh_092625.pdf)
- Programme has evolved over time into the current system but retains strong support.

## PROMs questionnaires are administered pre and post operatively to generate outcomes data.



- PROMs questionnaires comprise:
  - Two generic measures of Quality of Life: EQ-5D Index and EQ-VAS
  - One condition-specific measure of outcomes: Oxford Knee Score, Oxford Hip Score and Aberdeen Varicose Vein Score.
- Patients complete two questionnaires:
  - Pre-operative: completed by patients at a provider either on day of admission or at pre-operative assessment.
  - Post-operative: questionnaire is posted to patients at 3 months (Groin Hernia & Varicose Veins) or 6 months (Hip & Knee Replacement) after the procedure.
- Providers are supported to collect and report PROMs by a centrally appointed contractor to produce, distribute, collect and process PROMs data,
- Data collated and processed by Health and Social Care Information Centre (HSCIC) before routine publication as “Official Statistics”.

## The scale and scope of the PROMs programme is large and significant.

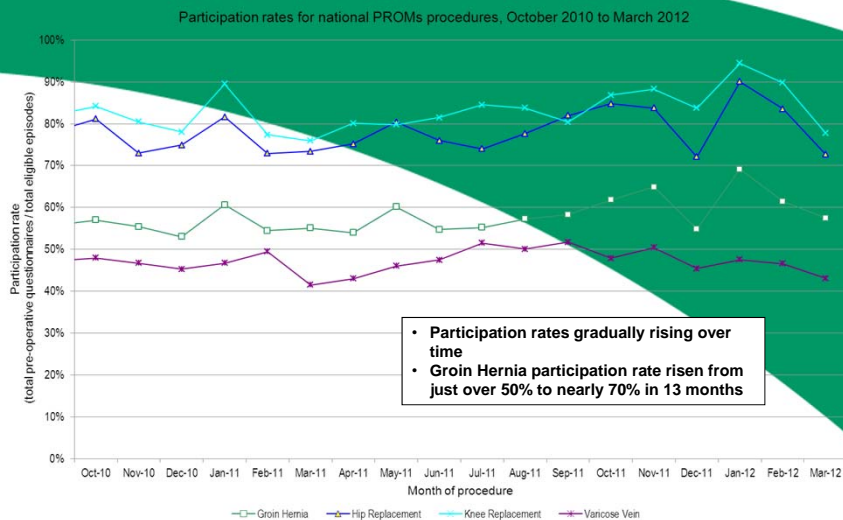


- PROMs national programme currently covers 4 elective procedures which cover around 250k patients per annum.
- Since the start of the programme in 2009 we have built possibly the most comprehensive dataset of patient reported outcomes in the world for these procedures:

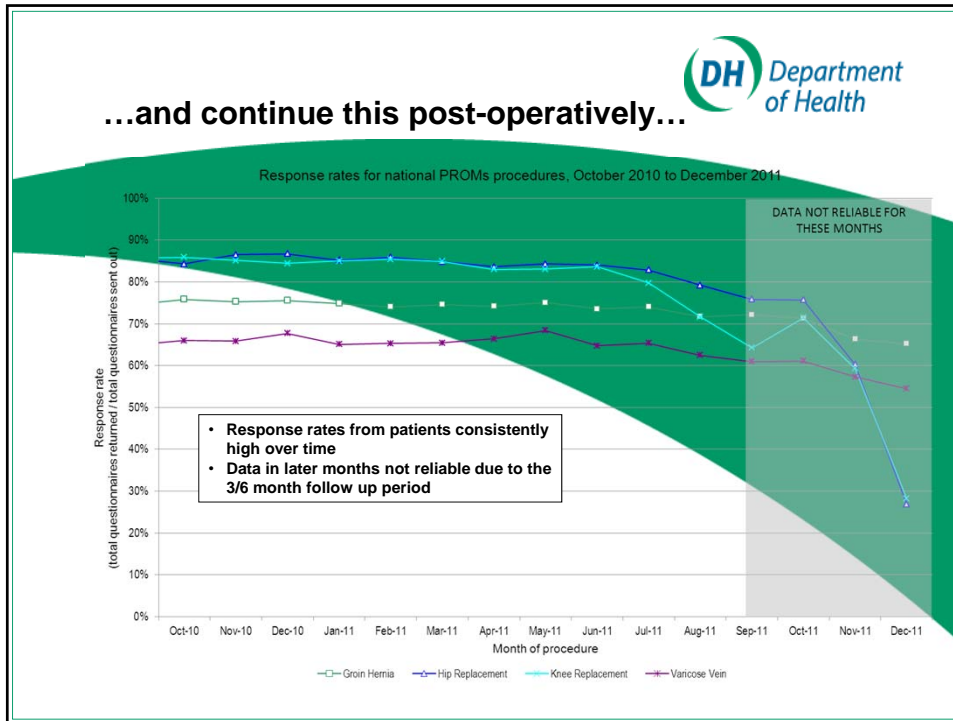
Procedure	Pre-operative questionnaires completed to date	Post-operative questionnaires returned to date
Groin Hernia Repair	106,000	69,000
Hip Replacement	146,000	99,000
Knee Replacement	169,000	112,000
Varicose Vein Surgery	40,000	23,000

- The initiative has generated significant interest from around the world with various attempts to replicate it.

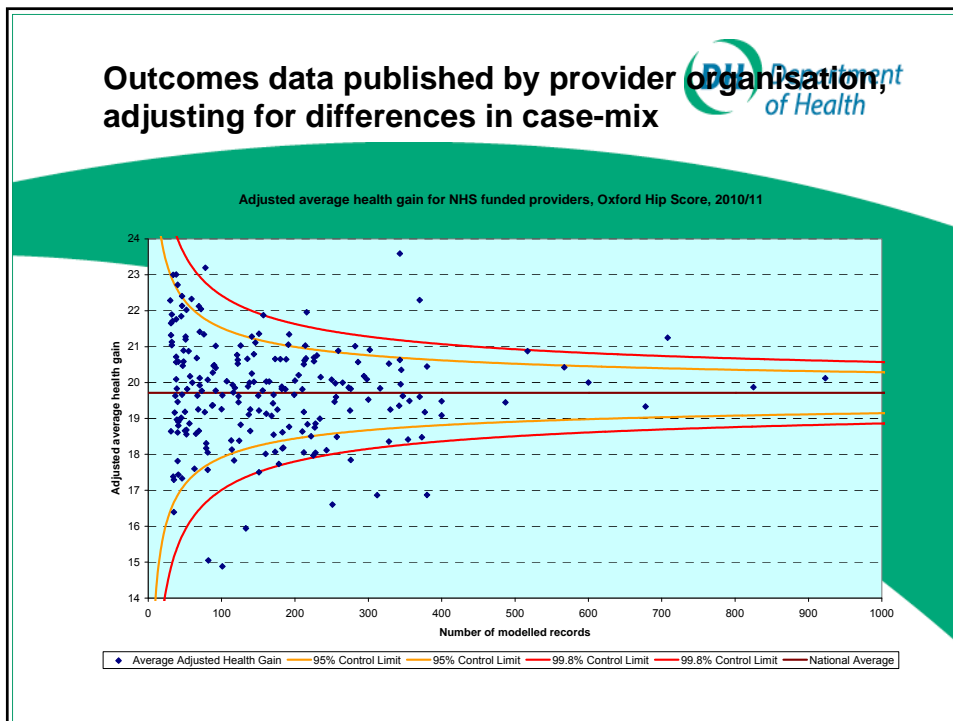
## We achieve and have maintained a high rate of patient buy-in especially for orthopaedics...



## ...and continue this post-operatively...



## Outcomes data published by provider organisation, adjusting for differences in case-mix



## These data have a range of potential applications...

**PROMs +**

Local use of data	» »	Benchmarking against peers leading to improvements
Quality Accounts	» »	Telling the story about performance
CQUIN	» »	Financial rewards for high quality; pay for better patient-reported quality...
Choice	» »	Choosing high quality providers
Resource Allocation	» »	Allocate scarce resources efficiently
Tackling health inequalities	» »	Appropriate access for given needs
Regulation (CQC)	» »	Assessing minimum standards
National accounting	» »	Driving economy-wide productivity improvements
Outcomes Framework	» »	Holding the NHS to account

## PROMs data are becoming part and parcel of the NHS Information landscape.

- Domain 2 and 3 of the NHS Outcomes Framework,
- Routinely published by Health and Social Care Information Centre, including identification of outlying units,
- Becoming a mandatory component of the Quality Accounts (2012/13),
- Incorporated into the CQC Quality and Risk Profiles (since Nov' 2011),
- Academic research,
- NHS Choices indicators,
- Local service improvement initiatives (myClinicalOutcomes.com, RNOH, metal on metal).

## We've recently highlighted some examples of where PROMs data are being used to change clinical practice:



- **Case Study 1: Improving Service Delivery**

Consultants at Royal Cornwall use PROMs data to monitor via a website post surgical patient health following hip and knee replacements. The initiative was set up in response to Royal Cornwall having poor scores in the early days of the PROMs programme. They hope online monitoring will improve patient health, encourage better compliance with post surgical therapy, e.g. physio, and reduce the need for face-to-face post-op outpatient appointments.

- **Case Study 2: Shaping Clinical Decisions**

Paul Baker *et al* published a paper in *The Journal of Bone & Joint Surgery* using OKS & EQ-5D PROMs data which investigates the relative effectiveness of Unicondylar Knee Replacements to Total Knee Replacements. They found both types of arthroplasty equally effective and questioned the use of UKR given the observed higher revision rates from worldwide registry data.

## Looking to the future



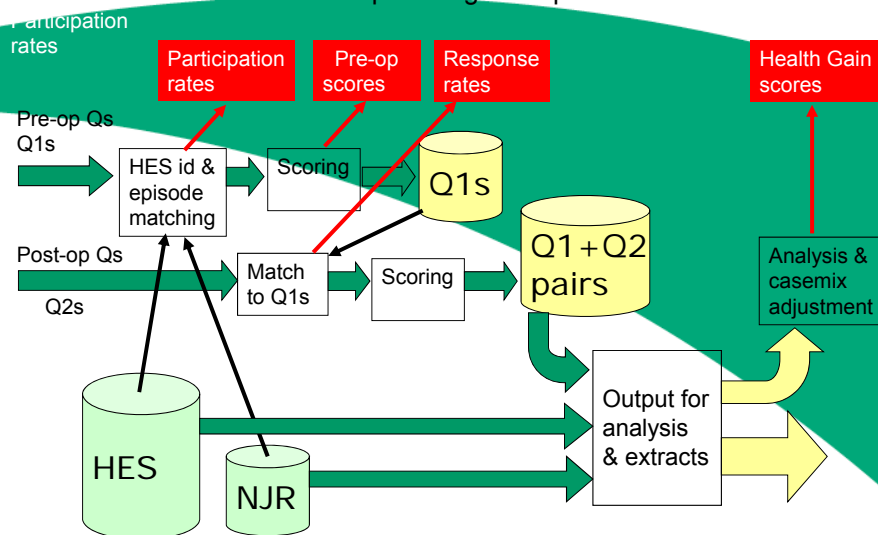
- There are a number of projects exploring PROMs in other clinical areas:
  - Elective Coronary Revascularisation (Coronary Artery Bypass Grafts and Angioplasty. Pilot began collecting data November 2011)
  - Cancer survivorship (Preliminary findings in – headline participation >60%)
  - Long-term conditions (due to report end of 2012)
  - Mental Health, Depression in secondary care (ethics approval awarded)
  - Pelvic Cancer (development work ongoing)
  - Musculoskeletal conditions (scoping work ongoing)
- There will also be work to identify areas of NHS activity in which to roll out a new shorter, sharper, generic PROMs Questionnaire.
- There will be more flexible data collection arrangements...
- ...and greater access to the patient level data for clinical teams.

## Summary

- Routine PROMs collection on a national scale gives us a unique insight into how effective patients feel their treatment has been. The amount of data collected and the high response rates make our findings robust.
- A large amount of work has gone into devising and developing the methods for presenting and analysing the data.
- PROMs has a huge variety of uses which we can use to drive quality improvement and incorporate patients' responses into their care.
- We are starting to see the evidence from where the data are being used to drive clinical practice.
- Work is under way to roll out PROMs into other clinical areas.

## National Implementation

How does it work? Example using Orthopaedic collection



# The Medicare Health Outcomes Survey (HOS)

*Liz Goldstein, Ph.D.*

*NQF PRO Workshop  
September 11, 2012*



## HOS Program Goal

- To gather valid, reliable, and clinically meaningful health status data in the Medicare Advantage (MA) program for use in quality improvement activities, plan accountability, public reporting, and improving health.



## Intended Uses for HOS Data

- Public reporting
- Pay for performance
- Quality improvement activities
- Program oversight
- Advance science of health outcomes research

19

## HOS Overview

- Implemented by CMS in 1998.
- All Medicare Advantage Organizations (MAOs) with >500 members must report.
- Annual data collection schedule consists of a baseline period and 2-year follow up period.
- Assesses an MAO's ability to maintain or improve the physical and mental health of its members.

20

# Survey Design and Questionnaire

- Sampling Unit: Medicare managed care contract
  - Less than 500 (exempt); 500 to 1,200 (all); greater than 1,200 (random sample)
- Longitudinal: Beneficiaries are re-surveyed in 2 years as part of the same cohort, if enrolled in the same plan at follow-up
- Administration: Self-administered, mailed survey with telephone follow-up
  - Surveys administered by NCQA-certified HOS vendors
  - 64 questions in total (HOS 2.0)

21

## Sample HOS Questions

1. In general, would you say your health is:

<b>Excellent</b>	<b>Very good</b>	<b>Good</b>	<b>Fair</b>	<b>Poor</b>
1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

2. The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

<b>ACTIVITIES</b>	<b>Yes, limited a lot</b>	<b>Yes, limited a little</b>	<b>No, not limited at all</b>
a. <b>Moderate activities</b> , such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.....	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
b. Climbing <b>several</b> flights of stairs.....	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>

3. During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health?**

a. <b>Accomplished less</b> than you would like.....	<b>No, none of the time</b>	<b>Yes, a little of the time</b>	<b>Yes, some of the time</b>	<b>Yes, most of the time</b>	<b>Yes, all of the time</b>
1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	
b. Were limited in the <b>kind</b> of work or other activities.....	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

[www.HOSonline.org](http://www.HOSonline.org)

22

## HOS Outcome Measure

- % whose physical and mental health improved over 2 years;
- % whose physical and mental health remained the same over 2 years; and
- % whose physical and mental health declined over 2 years.

*Two-year results are presented in terms of the percentages of beneficiaries who were better, the same, or worse than expected.*

23

## Reliability and Validity

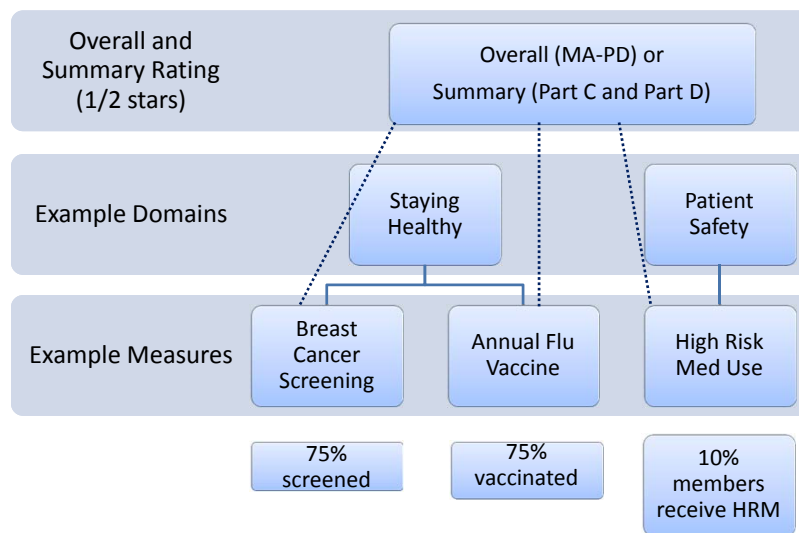
- The HOS was developed under the guidance of a Technical Expert Panel comprised of individuals with specific expertise in the health care industry and outcomes measurement.
- Continue to incorporate a theory and evidence-based approach to developing new methodology
- Continue to evaluate outcome measurement and case-mix methodology annually.

24

# Public Reporting

25

## Plan Ratings – Multiple Levels



26

## MA Plan Ratings Publicly Reported

Organization Name and Title						
Estimated Annual Drug Costs:[?]	Monthly Premium: [?]	Deductibles [?] and Drug Copay[?] / Coinsurance: [?]	Health Benefits: [?]	Drug Coverage [?], Drug Restrictions [?] and Other Programs:	Estimated Annual Health and Drug Costs: [?]	Overall Plan Rating:[?]
Retail Annual: \$0.00 Rest of 2012: \$0.00*	\$0.00 Drug: \$0.00 Health: \$0.00	Annual Drug Deductible: \$0 Health Plan Deductible: \$0 Drug Copay/Coinsurance: \$0 - \$85, 33%	Doctor Choice: Plan Doctors Only Out of Pocket Spending Limit: \$3,400 In-Network 	All Your Drugs on Formulary: N/A Drug Restrictions: N/A All Generics	\$2,050	★★★★ 4 out of 5 stars <input type="button" value="Enroll"/>

Overall Star Ratings

[www.medicare.gov](http://www.medicare.gov)



## Pay for Performance

## Medicare Advantage Quality Bonuses

- Required as part of the Affordable Care Act.
- CMS is conducting a demonstration to determine whether additional quality-based payments lead to more rapid and larger year-to-year quality improvements in Medicare Advantage (MA) plans' quality scores.
- Quality bonuses are based on the MA Plan Ratings.

29

## Quality Bonus Payments Under Current Law and CMS Demonstration

Quality Bonus %	Less than 3 stars	3 stars	3.5 stars	4/4.5 stars	5 stars
<i>Current law</i>	<i>none</i>	<i>none</i>	<i>none</i>	1.5-5%	1.5-5%
2012/2013 demonstration	none	3%	3.5%	4%	5%
2014 demonstration	none	3%	3.5%	5%	5%



## Summary of Uses of HOS Data

- CMS: assess performance of MA plans and reward high performing plans.
- MAOs and QIOs: monitor and improve health care quality.
- Medicare Beneficiaries: make better informed health care purchasing decisions.
- Health Researchers: advance the state-of-the-science in patient reported and functional health outcomes measurement and quality improvement interventions and strategies.

31



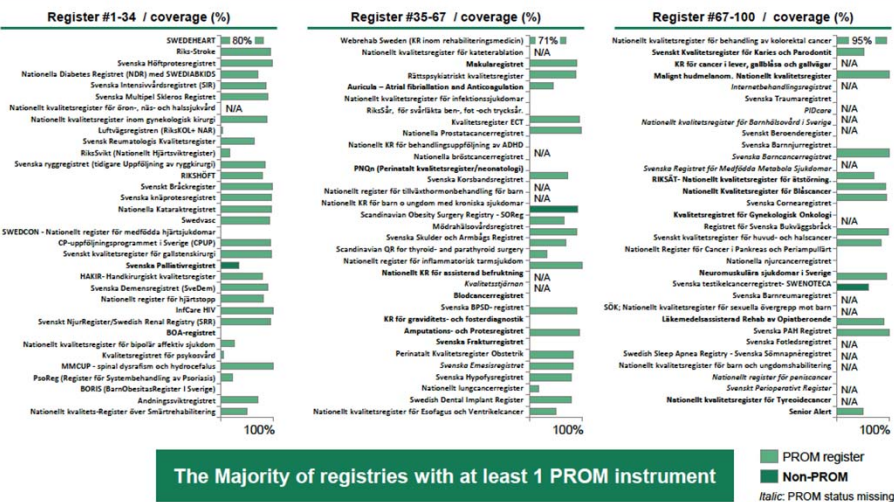
### **PROM use in Swedish healthcare outcomes registries**

September 11, 2012

THE BOSTON CONSULTING GROUP

32

## Sweden has 100 quality registers covering ~40% of HC spend A majority with >75% coverage of the relevant disease population



1. Defined as registered patients divided by total number of patients treated for register's symptom area(s)  
 Note: Data based on registers as of end of 2011, including "candidate" status registers. Source: SKI, PROMcentret, BCG analysis  
 PROMS\_Latest Version\_120907\_19h50\_PL\_E.pptx  
 THE BOSTON CONSULTING GROUP Draft—for discussion only 33 1

## Rationale for Sweden's growing use of PROM

1. The complexity of patients' health problems is increasingly understood
2. The number of available treatment options per condition have increased
3. Delivery systems are changing rapidly in response to increased economic and quality pressure, requiring better understanding of value of care provided
4. Individualized medicine calls for extensive research on the ways in which genetic, epigenetic, and other personal characteristics influence responses to therapy



## Specific funding by National payer association (SKL) to promote the use of PROMS in Swedish healthcare

### The Swedish PROM investment involve three competence centers

The Swedish PROM investment consist of a network of three cooperating competence centers mandated by SKL<sup>1</sup> to promote PROM in Sweden

The three organizations are governed by small steering committees and operated by small operating core teams of typically 5-15 people

- EyeNet's steering committee involve representatives from SKL and the Quality Registries and is operated by a small team of quality coordinators, IT consultants, epidemiologists and bio-statisticians

### The objective is to develop a more patient-centric healthcare

SKL has given the PROM network the mission:

- "to be a resource for the development of PROM within the national quality registries"

By fulfilling its objective:

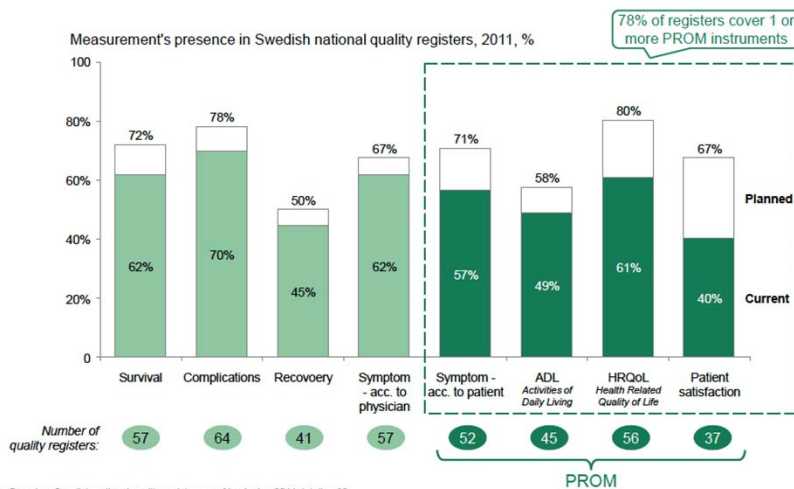
- "for Sweden to develop a more patient-centric healthcare through increased use of the results from PROM measurement for operational development work"

The network's primary day-to-day tasks involve to:

- Offer advice and support on the use and implementation of PROMS
- Hold seminars to inform on the best practice of PROMS, and what PROM to use in what situation
- Gather and publish publications related to PROM and its use in operational improvement work

<sup>1</sup> Swedish Association of Local Authorities and Regions  
Source: PROM Center, EyeNet Sweden and Registercentrum Västra Götaland  
PROMS\_Latest Version\_120907\_19h50\_PL\_E.pptx

## The 4 major PROM dimensions are covered in ~40-60% of Sweden's ~100 national quality registers



Note: Based on Swedish national quality registers as of beginning 2011, totaling 92  
Source: PROMcenter, BCG analysis  
PROMS\_Latest Version\_120907\_19h50\_PL\_E.pptx

## Overview of Swedish national quality registers (I)

Condition/area	# of registers	% of registers including instrument:					
		Symptom	ADL	HRQoL			Satisfaction
				EQ-5D	SF-36	Total	
Cancer	15	33%	27%	13%	0%	27%	33%
Locomotive organs	14	71%	71%	64%	36%	79%	64%
Obstetrics and children's diseases	10	40%	40%	20%	0%	60%	30%
Psychiatry	9	40%	40%	10%	0%	40%	40%
Circulatory organs	7	71%	57%	57%	0%	71%	29%
Eyes, ears, dental	7	71%	43%	0%	0%	57%	57%

Note: Measures not specified include SHS, GAF, PedsQL, CHAQ, Disabkids, Oxford-12, WOOS, FAOS, QuickDAS & PDQ8; Based on Swedish national quality registers as of beginning 2011, totalling 52  
 Source: PROMcenter, BCG analysis  
 PROMS\_Latest Version\_120907\_19h50\_PL\_EE.pptx

## Overview of Swedish national quality registers (II)

Condition/area	# of registers	% of registers including instrument:					
		Symptom	ADL	HRQoL			Satisfaction
				EQ-5D	SF-36	Total	
Endocrine and skin diseases	6	86%	57%	29%	14%	71%	57%
Nervous system	6	50%	67%	50%	0%	50%	50%
Stomach / bowel	5	80%	20%	20%	40%	80%	20%
Palliative and age related	5	20%	40%	20%	0%	40%	0%
Emergency treatment	2	33%	67%	0%	33%	33%	0%
Other	6	83%	33%	17%	33%	83%	33%

Note: Measures not specified include SHS, GAF, PedsQL, CHAQ, Disabkids, Oxford-12, WOOS, FAOS, QuickDAS & PDQ8; Based on Swedish national quality registers as of beginning 2011, totalling 52  
 Source: PROMcenter, BCG analysis  
 PROMS\_Latest Version\_120907\_19h50\_PL\_EE.pptx

## International Consortium for Health Outcomes measurements (ICHOM) bringing together international PROM tools

PROMs	Acute non-surgical	Chronic surgical					Chronic non-surgical		
	Australian Stroke Clinical Registry	Swedish Spine Register (Swespine)	Singapore Gen. Hosp. Low Back Pain Registry	Norwegian Arthroplasty Registry	Swedish Hip Arthroplasty Register	California Joint Rep. Registry	New Zealand Joint Registry	Schoen Klinik Depression Registry	Swedish COPD Registry (Rikskol)
<b>Generic</b>									
SF-36		◆	◆						
RAND 36			◆						
SF-12									
EQ-5D	◆	◆	◆	◆	◆				
WOMAC									
<b>Condition-specific</b>									
Oswestry		◆	◆						
Oxford 12							◆		
BSI-1								◆	
PHQ-D1								◆	
CCQ									◆

EQ-5D most widely used PROM across registries

The use of multiple PROM tools may capture greater outcome details

PROMS\_Latest Version\_120907\_19h50\_PL\_E.pptx

THE BOSTON CONSULTING GROUP

Draft—for discussion only

39 7

Copyright © 2012 by The Boston Consulting Group, Inc. All rights reserved.

## PROM in Swedish hip arthroplasty has enabled quality and efficiency improvements

### Significant efficiency gains from PRO metrics in THA

Main indications for total hip arthroplasty are pain and impaired HRQoL, making PRO key

Hence, a standardized PROM protocol was introduced in 2002

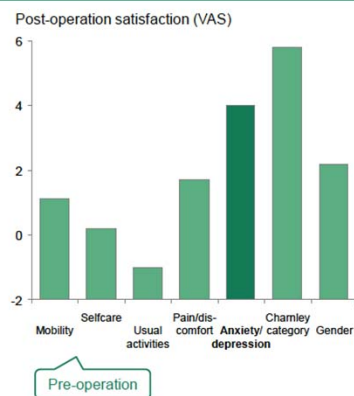
- Including Charnley functional categories, VAS, EQ-5D instruments and EuroQol
- Covering all but one Swedish THA performers

~8 kUSD cost per THR patient, of which 61% in productivity loss, and >16k patients annually, implies significant QoL and cost gains potential

Based on PROM output, two alternative treatment procedures could be properly assessed

- One-stage bilateral THA came out on top, with better HRQoL profile, at lower cost

### Pre-operative anxiety/depression stands out as deciding factor in THA satisfaction



Source: Patient-reported Outcome Measures and Health-economic Aspects of Total Hip Arthroplasty, Rollfson (2010); BCG analysis

PROMS\_Latest Version\_120907\_19h50\_PL\_E.pptx

THE BOSTON CONSULTING GROUP

Draft—for discussion only

40 8

Copyright © 2012 by The Boston Consulting Group, Inc. All rights reserved.

## PROMs in SRQ has transformed treatment of rheumatoid arthritis in Sweden



### SRQ with high PROM content became a role model for patient involvement

Swedish Rheumatology Quality registry (SRQ) has been in place since 1996 and containing PROMs since 2004

PROMs are used in SRQ to create multi-dimensional follow-ups of medical and functional quality from the patient perspective

- The patient can register his/her own assessment of current symptoms, day-to-day functionality and quality of life via the Internet
- The patient input is then translated to quantifiable measures and instantly reported to the patient
- The results are openly shared and available for the profession, hospital management, patients and the public

SRQ is a national role model for patient involvement

- SRQ's success laid the foundation for project KUR<sup>1</sup>, aiming to leverage PROM's success in SRQ in other quality registers and increase cooperation between industry, health care and academia

### PROM with several medical and organizational benefits

The inclusion of PROM helped create a more holistic measurement of outcome of treatments

- PROM enabled identification of best practice treatments of a diversified patient population containing old and multi-ill patients

PROM increased the patients' knowledge and influence over treatments and enhanced communication quality between patient and physician

- Patient involvement and transparency helped the patient to monitor his/her health and treatment effectiveness over time

PROM increased the status of subjective dimensions

- PROM helped patient and physicians to quantify "subjective" dimensions

PROM lead to improved resource allocation, capacity planning and treatment level in diseases with high variation in disease activity, e.g. RA

- E.g. Karolinska in Stockholm identifies patients with high and low disease activity and can plan treatments and resources accordingly and instantly

1. Klinisk Utveckling via Register, initiated by the Swedish Enterprise Ministry  
Source: Svensk Reumatologisk Kvalitetsregister – ger kvalitetsförbättring i den kliniska vardagen. Årsrapport 2009-10. Lindblad et al (2010)

PROMS\_Latest Version\_120907\_19h50\_P1\_E.pptx

## Recap of Key Characteristics for Selecting PROMs for Use in Performance Measurement

Karen Adams, NQF



NATIONAL  
QUALITY FORUM

### Terminology

- PRO measure (PROM): Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).
- PRO-based performance measure (PRO-PM): A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

## Key Characteristics for Selecting PROMs for Use in Performance Measurement

Workshop 1

- Expert Panel discussed high leverage characteristics for identifying PROMs most suitable for development and testing of performance measures.

Workshop 1

- Expert Panel agreed the psychometric properties detailed in the commissioned paper were considered as baseline—but also offered additional guideposts for consideration.

Pre-work for  
Workshop 2

- NQF staff drafted language to capture these additional characteristics.
- Expert Panel provided edits and feedback via a survey.

Workshop 2

- Further refinement of characteristics and their relationship to NQF endorsement criteria.

## Actionability: Key themes from survey


- Key end users including patients, providers and systems should be motivated by the PROM to lead to improvement.
- Evidence should indicate that care can be improved in a relatively short time period for patient respondents.
- Should take into consideration a range of evidence including expert consensus not only RCT level.
- Some outcomes are worth measuring that may not be amendable to change by providers—but patients need to make informed decisions (e.g., pain after intervention, functional status after treatment)

## Meaningfulness: Key themes from survey

- Burden to respondents and administrators has implications in determining meaningfulness.
- Important concepts for PROMs to capture include the patient's perspective on the impact of the condition and/or treatment on the individual's life.
- Need to also capture long-term care services and supports beyond acute or episodic care.
- Caregivers perspectives should also be included in determining meaningfulness.

## Facilitate shared decision-making: Key themes from survey

- Some indecision around the capacity for facilitating shared decision-making (SDM) being a key characteristic.
  - Not all performance measures need to facilitate SDM.
  - Not all patients want SDM and these are not uniformly distributed.
  - Redundant with actionability.
- Decoupling shared decision-making from aggregation issues
  - PROM being sufficiently standardized to permit aggregation or roll up to a population or accountable entity.
- Broaden to patient engagement –SDM a process and too narrowly defines engagement.



## Implementable: Key themes from survey

- Definition combines many elements –hard to map to or meet all requirements
  - Covered under actionability (amenable to change) and meaningfulness (usability).
  - Measures of disparity not indicators of implementability.
- Ease of fielding is an important consideration and its relationship to burden.



## Overview of NQF Criteria

Karen Pace, NQF



NATIONAL  
QUALITY FORUM

Recap of Key Characteristics  
for Selecting PROMs for Use  
in Performance Measurement

Overview of NQF Criteria



NATIONAL  
QUALITY FORUM

## Terminology

- **Patient-reported outcome (PRO):** The concept of any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.
- **PRO measure (PROM):** Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).
- **PRO-based performance measure (PRO-PM):** A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

## Characteristics for Selecting PROMs - Psychometric Properties

### NQF Criteria

- Scientific Acceptability of Measure Properties
  - **Reliability**
  - **Validity**
- NQF endorses PRO-PM, not the PROM
- PROM value – data used in performance measure
- NQF criteria for the performance measure allows for testing of reliability and validity of data elements or the performance measure score
- Will discuss reliability and validity of the performance measure score in the methods panels that follow

## Additional Characteristics for Selecting PROMs - Actionability

### NQF Criteria - Importance to Measure and Report

- Performance gap/Opportunity for improvement
  - Data demonstrating considerable variation in performance  
OR overall less than optimal performance
  - Data on disparities in care
- Evidence to Support the Measure Focus
  - The measure focus is a health outcome **OR**
  - Evidence-based – evidence linking structure, process, or intermediate clinical outcome to desired health outcome
    - Quantity, quality, consistency of body of evidence

## NQF Criteria - Evidence

- Health outcome:<sup>3</sup> a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,<sup>4</sup> or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence<sup>5</sup> that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.

## Additional Characteristics for Selecting PROMs - Meaningfulness

### **NQF Criteria - Importance to Measure and Report**

- Evidence - The measure focus is a health outcome or is evidence-based
  - Experience with care - valued by patients; patient is the best source of information; or linked to desired outcomes
- High impact
  - National health goal or priority
  - Data on numbers of persons affected, high resource use, severity of illness, consequences of poor quality
- Usability and Use

## NQF Criteria - Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement<sup>2</sup> to achieve the goal of high-quality, efficient healthcare for individuals or populations.

- **Accountability and Transparency**
  - In use for public reporting or other accountability application
  - If not, is there a credible plan and progress?
- **Improvement**
  - Progress in achieving high-quality, efficient healthcare
  - If not in use at initial endorsement, is there a credible rationale for how the measure could further goal of high-quality, efficient healthcare

## Additional Characteristics for Selecting PROMs - Implementable

### NQF Criteria

- Feasibility
  - Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.

## NQF Criteria - Feasibility

- **Clinical data generated and used during care process**
  - E.g., blood pressure, lab value vs. survey or observation
- **Electronic sources**
  - EHR, claims vs. abstracted and entered into database/registry
  - Is there a credible, near-term path to electronic collection?
- **Data collection strategy can be implemented**
  - Is it already in operational use or testing indicated ready for operational use?

## Methods that Contribute to Trust - Demonstrating Reliability of PRO-PMs

### Overview of NQF Criteria



NATIONAL  
QUALITY FORUM

## Terminology

- **PRO measure (PROM):** Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).
- **PRO-based performance measure (PRO-PM):** A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

## NQF Criteria - Scientific Acceptability of Measure Properties

### Reliability

- **2a1.** The measure is well defined and precisely specified<sup>7</sup> so it can be implemented consistently within and across organizations and allow for comparability. EHR measure specifications are based on the quality data model (QDM).<sup>8</sup>
- **2a2.** Reliability testing<sup>9</sup> demonstrates the measure **data elements** are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the **measure score** is precise.

## NQF Measure Testing Guidance

- **Empirical analysis** to demonstrate the reliability and validity of the **measure as specified**
- Allow for testing at one level – data element (PROM value) or performance measure score
- Appropriate method & scope with acceptable results
- Reliability of **data elements** refers to repeatability and reproducibility of the data elements for the same population in the same time period.
- Reliability of the **measure score** refers to the proportion of variation in the performance scores due to systematic differences across the measured entities (signal) in relation to random variation or noise.

## Methods that Contribute to Trust - Demonstrating Validity of PRO-PMs, Part 1

### Overview of NQF Criteria



NATIONAL  
QUALITY FORUM

## Terminology

- **PRO measure (PROM):** Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).
- **PRO-based performance measure (PRO-PM):** A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).



## NQF Criteria - Scientific Acceptability of Measure Properties

### Validity

- **2b1.** The measure specifications<sup>7</sup> are consistent with the evidence presented to support the focus of measurement. The measure is specified to capture the most inclusive target population indicated by the evidence, and exclusions are supported by the evidence.
  
- **2b2.** Validity testing<sup>10</sup> 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.

## NQF Measure Testing Guidance

- **Empirical analysis** to demonstrate the reliability and validity of the **measure as specified**
- Allow for testing at one level – data element (PROM value) or performance measure score
  - Systematic assessment of face validity of measure score
- Appropriate method & scope with acceptable results
  
- Validity of **data elements** refers to the correctness of the data elements as compared to an authoritative source.
- Validity of the **measure score** refers to 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).

## Methods that Contribute to Trust - Demonstrating Validity of PRO-PMs, Part 2

### Overview of NQF Criteria



NATIONAL  
QUALITY FORUM

## Terminology

- **PRO measure (PROM):** Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).
- **PRO-based performance measure (PRO-PM):** A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

## Threats to Validity

- Conceptual – measure focus is not a relevant outcome of healthcare or not strongly linked to a relevant outcome
- Unreliability – generally an unreliable measure cannot be valid
- Patients inappropriately excluded from measurement
- Differences in patient mix for outcome and resource use measures not adjusted
- Measure scores generated with multiple data sources/methods
- Systematic missing or “incorrect” data (unintentional or intentional)

## NQF Criteria - Scientific Acceptability of Measure Properties

### Validity

- **2b3. Exclusions** are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;<sup>11</sup>

### AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).<sup>12</sup>

## NQF Criteria - Validity

- **2b4.** For **outcome** measures: an evidence-based **risk-adjustment** strategy (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome\* (but not factors related to disparities in care or the quality of care) and are present at start of care;<sup>13,14</sup> and has demonstrated adequate discrimination and calibration **OR** rationale/data support no risk adjustment/ stratification.

\* Could include baseline status (baseline PROM value)

## NQF Criteria - Validity

- **2b5.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful<sup>15</sup> **differences in performance**; **OR** there is evidence of overall less-than-optimal performance.
- **2b6.** If **multiple data sources/methods** are specified, there is demonstration they produce comparable results.

## Measure Testing Guidance

- Empirical analysis to demonstrate the reliability and validity of the *measure as specified* including analysis of **issues that pose threats to the validity of conclusions about quality** of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.
- Appropriate method & scope with acceptable results

## NQF Measure Evaluation Guidance

- Measure Evaluation Criteria
- Reports on guidance for measure evaluation:
  - Evidence for the Focus of Measurement and Importance to Measure and Report
  - Measure Testing and Scientific Acceptability of Measure Properties
  - Usability and Use
  - Measure Harmonization
  - Competing Measures
  - Guidance on Measure Construction

# Methods that Contribute to Trust – Demonstrating Reliability of PRO-PMs

Laura Smith, PhD  
Anne Deutsch, RN, PhD, CRRN



NATIONAL  
QUALITY FORUM

## Reliability

- Reliability is a necessary, though not sufficient, pre-condition for validity
- Lack of reliability in performance measures can result in misclassification of providers in quality rankings

## Patient-Level and Performance Measure-Level Reliability

- Patient-level:

$$\text{reliability} = \frac{\text{subject variability}}{(\text{subject variability} + \text{measurement error})}$$

- Performance Measure-level:

$$\text{reliability} = \frac{\text{signal}}{(\text{signal} + \text{noise})}$$

- Range is 0-1
- Thresholds for performance measure comparisons:
  - among groups: 0.70
  - for individuals: 0.90

## Determinants of Performance Measure Reliability

- Magnitude of true differences among providers
- Within provider-variation
- Provider sample, or denominator, size

## Determinants of Performance Measure Reliability

- What are the implications of various approaches to aggregating PROM data (e.g., average/median amount of change; percentage who improve/reach benchmark/ have meaningful change) on reliability of the PRO-PM score?

## Determinants of Performance Measure Reliability

- Performance measure reliability is dependent on the characteristics of the set of providers and patients included in the measure specifications
- Reliability is not static
- Estimates for smaller providers are more vulnerable to random error



## Performance Measure Reliability Testing Methods

- What methods for reliability testing would support the demonstration of reliability of the PRO-PM scores?
- Are there any differences or unique considerations for demonstrating and evaluating the reliability of a PRO-PM (as compared to other quality performance measures)? Is reliability of the PRO-PM score needed in addition to reliability of the PROM?

## Performance Measure Reliability Testing Methods

- Two-level hierarchical models (to estimate signal and noise)
- Examine overlap in confidence intervals calculated for each provider's PRO-PM score
- "Interunit" reliability (derived from F-test)
- Intra-class correlation coefficient (mean patient values within provider)
- Generalizability theory (factorial analysis of variance)
- Monte-Carlo simulation

## Provider Size and Performance Measure Reliability

- What is the likelihood of providers achieving an adequate sample size for a specific level of analysis (e.g., clinician or hospital) to provide a reliable estimate?
- What are the minimum sample sizes needed to reliably calculate provider-level (individual provider or organization) scores on PRO-based performance measures?

## Provider Size and Performance Measure Reliability

- Signal to noise calculation of reliability derived from hierarchical modeling can allow some estimation of minimum sample size needed
- Examining the relationship between provider reliability estimates and their measure denominator sizes can be used to identify the threshold size where reliability estimates are 0.70 or greater

## Strategies for Improving Performance Measure Reliability

- Design composites (combining performance measures) to increase data points being used
- Increase provider sample size or denominator by having measure include more calendar time
- Improve reliability of underlying PROM
- Apply reliability adjustment

## Reliability Adjustment of Performance Measures

- Shrink provider PRO-PM estimates towards the mean value for
  - All providers
  - Providers of similar size or volume
  - Providers of similar size or volume, and other characteristics
- Shrinkage is greater for smaller providers because their estimates are more vulnerable to random error

## Relationship Between Performance Level Reliability and Validity

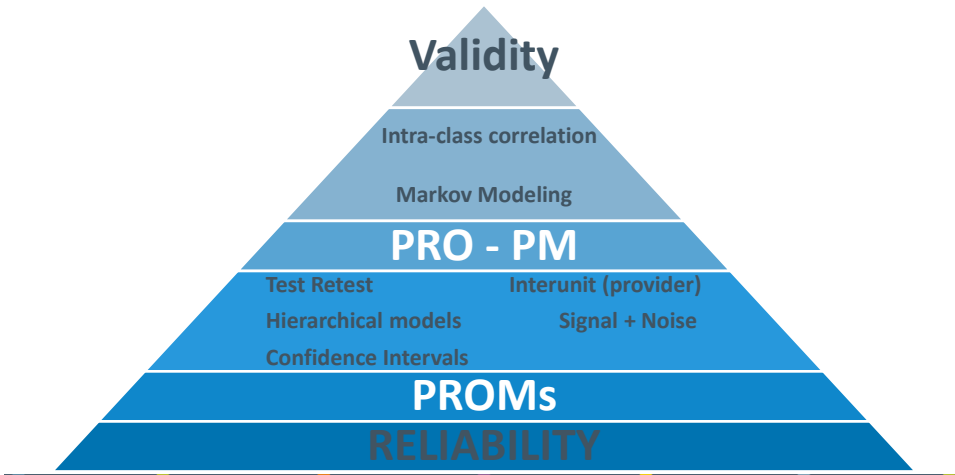
- What impact does poor reliability of the PRO-PM score have on validity of the PRO-PM as an indicator of quality?

# Methods that Contribute to Trust – Demonstrating Reliability of PRO-PMs

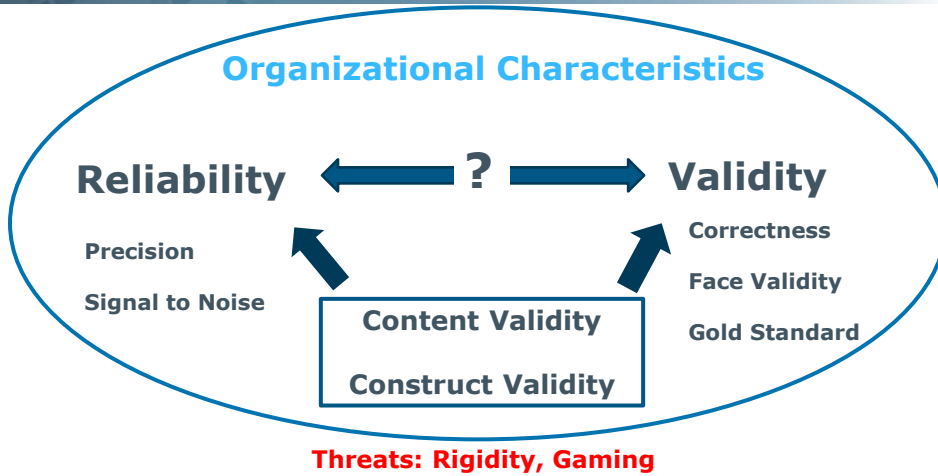
Lewis Kazis, Boston University School of Public Health



NATIONAL QUALITY FORUM



# Interaction between Reliability and Validity



# Methods that Contribute to Trust – Demonstrating Validity of PRO-PMs, Part 1

Anne Deutsch, RN, PhD, CRRN  
Barbara Gage, PhD



NATIONAL  
QUALITY FORUM

## Definitions

**PRO patient-level measure/instrument (PROM):** *Tools* to assess health condition at the individual level (e.g., health status and status of physical, mental, and social functioning, health behavior, or experience with healthcare).

**Performance measure:** Numeric quantification of healthcare quality for a designated accountable healthcare entity, such as a hospital, health plan, nursing home, clinician, etc.

**PRO-based performance measure (PRO-PM):** An organizational performance measure based on patient-reported outcome data aggregated for an accountable healthcare entity (e.g., percentage of patients in an organization whose depression score as measured by the PHQ-9 improved).

## Aggregating PROM data

- What are the implications of various approaches to aggregating PROM data (e.g., average/median amount of change; percentage who improve/reach benchmark/ have meaningful change) on:
  - the validity of conclusions about quality; and
  - the ability to discriminate performance among accountable entities?

## Change vs. Threshold Value

Two options for reporting the provider-level outcome or health status:

- 1) a change in status (e.g., decrease in pain between start of care and end of care, increase in functional status between start and end of care)
- 2) a threshold achieved (e.g., percent of patients with moderate to severe pain)



## Change Scores

- Individuals' change scores can vary in magnitude and direction of change; at provider level, individual differences could be masked
- Change scores tend to have lower reliability than baseline and follow-up scores
- Floor and ceiling effects: real changes in health status occur, but instrument not sensitive
- Clinical meaning of change score may not be known

## Examples

- "Change in basic mobility as measured by the AM-PAC": change measure to document improvement in functional status
- "Percent of patients with moderate to severe pain": threshold value
- "Depression remission within 6 Months": threshold value, but reflects a change from a baseline PHQ-9 score (possible depression) to a follow-up score (no depression)

## Performance Score

- Provider-level calculation: mean or median, percent or ratio
- For continuous data, mean or median can be calculated
- Mean or median may not represent diversity of patients if population is heterogeneous or data not normally distributed at the provider level

## Performance Score: Percent

- Calculate a percentage value based on the number of patients who achieve or exceed a benchmark:
  - 1) a national expected value (threshold or change) based on outcomes of similar patients
  - 2) a fixed amount of change defined based on a PROM-specific clinically important difference or PROM-specific minimal detectable change
  - 3) a threshold value that is associated with a longer-term outcome (e.g., balance score associated with a reduced risk of falls).
- For PROMs that have established clinically meaningful thresholds (i.e., cut points), the performance score should incorporate these thresholds

## Performance Score: Ratio

- Ratio: a score that may have a value of zero or greater that is derived by dividing a count of one type of data by a count of another type of data (e.g., the number of patients reporting pain score of 7 or higher divided by the number of inpatient days)
- A ratio may be preferred when the amount of time (e.g., number of days) that a patient is at risk for the outcome is important

## Examples

- “Depression Remission within 6 Months”: classifies scores into clinically meaningful groups ( $< 5$  = not depressed and  $> 9$  = depressive symptoms), and the patient is considered to have made an improvement if he or she moves from the depressive symptom category ( $> 9$ ) to the not depressed category ( $< 5$ )
- “Change in basic mobility as measured by the AM-PAC”: percent of patients with a change/improvement between admission and discharge. Change is defined as a difference of one or more minimal detectable change(s)
- A minimal detectable change refers to the minimal amount of change that is not likely to be due to measurement error, and thus represents a true change

## Validity Testing

- What methods for validity testing would support the demonstration of validity of the performance measure score for making conclusions about quality of care?
- Are there any differences or unique considerations for demonstrating and evaluating the validity of PRO-PMs (as compared to other quality performance measures)?

## Face Validity

Face validity:

- modified Delphi survey
- formal consensus process
- UCLA/RAND Appropriateness Method
- American College of Cardiology and American Heart Association Methodology for the Selection and Creation of Performance Measures

## Face Validity: Patient Experts

- Face validity of PRO-PMs could also be tested with “patient experts” using qualitative research methods, such as focus groups, semi-structured interviews, and cognitive interviews
- If patient experts are used, it will be critical to describe and frame the concept of healthcare quality. Hibbard\* provides a foundation for this framing

\*Hibbard JH, Greene J, Daniel D. What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Medical Care Research & Review*. 2010;67:275-293

## Criterion Validity

- Extent that the measure agrees with a “gold standard”
- For a PRO-PM, comparison of a performance score based on clinician observation that taps into the same construct (e.g., functional status) may be one way to demonstrate concurrent validity

## Construct Validity

- How the measure performs based on theory
- Example: Identify providers who implemented quality improvement (QI) initiatives focused on a PRO construct and compare the providers' performance scores before and after the QI program

## Validity of Performance Score

- Is validity of the performance score as an indicator of quality needed in addition to validity of PROM?

## Methods that Contribute to Trust – Demonstrating Validity of PRO-PMs, Part 2

Anne Deutsch, RN, PhD, CRRN  
Barbara Gage, PhD



NATIONAL  
QUALITY FORUM

### Risk Adjustment

- Are there any differences or unique considerations for risk adjustment of a PRO-PM (as compared to other quality outcome performance measures)?

## Risk Adjustment: Covariates

- Patient factors selected for risk adjustment of a PRO-PM should be based on evidence that the factor affects the outcome
- Evidence: peer-reviewed research literature, clinical expert opinion. Informed patients could provide very valuable insights into potential covariates
- Covariates: different for different PRO concepts

## Potential Covariates

- Patient demographic factors: age
- Patient clinical factors that are present at the start of care: diagnosis, severity of illness, comorbidities, and baseline scores
- Psychosocial factors: adherence, motivation, understanding, engagement, and readiness to change, have been suggested as potential covariates for PRO-PMs



## Risk Adjustment Methodology

- Stratify by risk groups (strata)
- Regression model
- Strata and regression modeling
  
- Regression modeling issue: hierarchical generalized linear models (HGLMs) vs. fixed-effects regression models

## Incomplete/Missing Data

- What are the implications of exclusions, incomplete/missing data, and response rate/bias on validity of the performance measure and the testing needed to assess impact on validity?

## Reasons for Missing Data

- Self-administration: non-response
- Interviewer-administered: not conducted by clinician
  
- Person unable to respond due to cognitive limitations, young age, language barriers, etc.

## PRO-PM Testing

- Response rates for proposed PRO-PMs should be reported as part of the testing results
- PRO-PM description should describe the mode of administration
- PRO-PM description should address use of proxy responses and method of data collection

## Example

- “Percent of residents with moderate to severe pain (short stay)”: data are collected using an interview as part of the mandated Minimum Data Set. This has resulted in relatively low missing data rates

## Use of Proxies

- What are the implications of using proxies on the validity of the performance measure and the testing needed to assess impact on validity?

## Proxies

- In order to use proxy responses within a PRO-PM, proxy responses would need to be reasonably accurate
- Proxies have demonstrated acceptable reliability for PROs such as functional status, where the proxy can observe the patient
- Proxy responses are less useful for more subjective PRO concepts, such as pain intensity, nausea, depression symptoms, because proxy data in this area tend to be less reliable

## Proxies (cont'd)

- Proxy responses are reasonable to consider for child health measures where parents are proxies and the research has show small differences in child-parent reports
- Use of proxies may minimize missing data, but it may introduce error to the performance score, and thus would be a threat to validity

## PRO-PM with multiple PROMs

- What are the implications of specifying more than one PROM (i.e., instrument/scale) in a performance measure and the testing needed to assess impact on validity?

## PRO-PM with multiple PROMs (cont'd)

- Use of different PROMs to measure the same construct: research demonstrating the agreement with assignment to clinically important groups (e.g., depressed, not depressed) should be high
- If assignment into clinically meaningful groups is not well aligned, this may introduce systematic errors based on the instruments selected

## Example

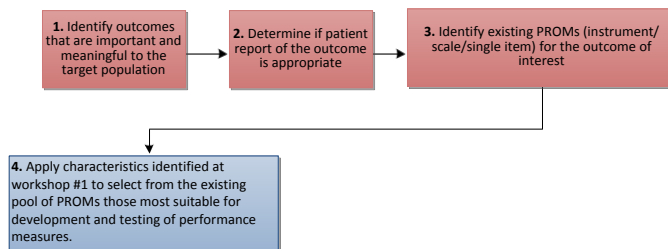
- “Percent of residents with moderate to severe pain (short stay)” allows for pain data to be collected based on the numeric rating scale (0 to 10 scale) or the pain verbal descriptor scale (mild, moderate, severe, very severe/horrible)
- The PRO-PM equates thresholds of pain across the 2 items

## Revisit pathway from individual-level PROM to NQF-endorsed PRO-PM

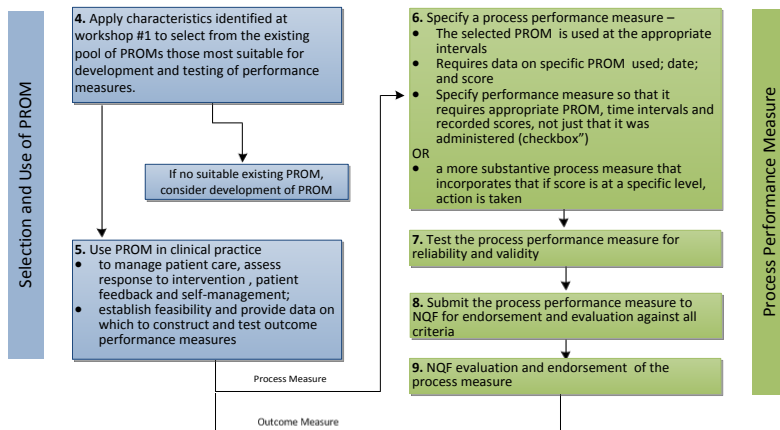


NATIONAL  
QUALITY FORUM

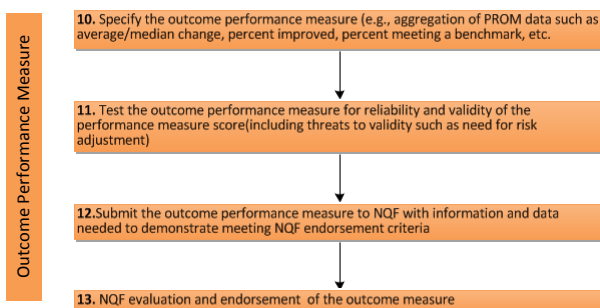
### Pathway



# Pathway



# Pathway







## Pathway

- [PRO Day1\\_am0900\\_Day2\\_am1050\\_Pathway.pdf](#)