Going to the PROM: Patient Reported Outcome Measures

Laurie Gutmann, MD

OUTLINE

- Background—why PROM's are important
  - The role of the patient in science
  - FDA regulatory guidance
  - Role of PROM's in how we'll get paid in future for clinical care….
- Science—how to develop a PROM
  - Development
  - Validation
- Examples from real life

Rationale and Definition

- Patient perspective through patient reported outcome measures (PROMs)
  - Crucial element for clinical care, quality performance management and clinical research.
- Direct patient report regarding their health condition and treatment
  - Symptoms
  - Functional status
  - Health-related quality of life.
- Types of PROMs
  - Generic -- appropriate for use in a wide range of conditions
  - Specific -- focus on the specific symptoms and side effects of a given disease, condition or treatment.

What a PROM is and isn't

- **NOT** a proxy reported measurement, by definition
- **NOT** an observer reported measurement, although can be combined with other data

- **IS** one way to get information useful for clinical care or research
  - Useful in combination with other non-PROM measures
- **IS** a report of symptoms or health state ,and/or a measure of QOL
  - Based on a theoretical framework, multiple domains typical

Role of the Patient

- In clinical care
  - PROM can be one way for a patient to quickly communicate important information (time of clinical visits are short)
- In research:
  - Can provide information that is otherwise hard to obtain (example = spasticity hard to measure/characterize)
  - Can be obtained via EHR or in written form
- HRQOL not routinely assessed, but really matters!
  - A valid and meaningful endpoint for research
  - Can give us important clinical perspective, tied to patient satisfaction measures

Factors affecting HRQOL

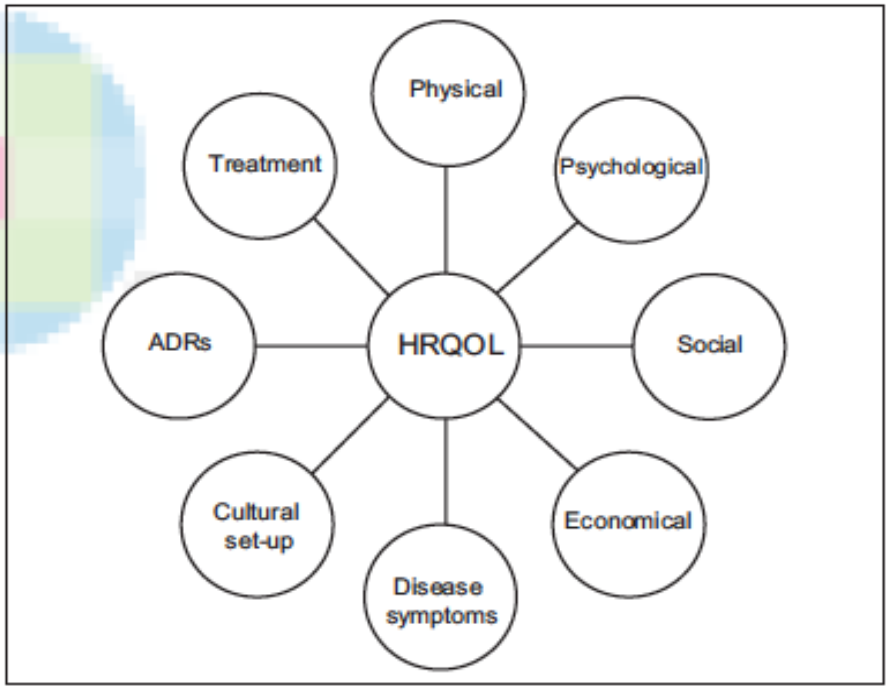


Figure 3: Factors affecting health related quality of life

Examples of PROM's

- Symptoms of depression: PHQ-9, PHQ-2
- HRQOL: EQ-5D, SF-12 or SF-36 (generic), SSQOL (disease specific)
- Patient-Reported Outcomes Measurement Information System (PROMIS)
  - In late 2004 the NIH initiated a multi-center cooperative group - PROMIS
  - Goal: to build and validate common, accessible item banks to measure key symptoms and health concepts applicable to a range of chronic conditions to enable efficient and interpretable clinical trial and clinical practice applications of patient-reported outcomes (PROMs).
- NeuroQOL, NIH Toolbox
- Computerized adaptive testing (CAT) tools for tailored individual assessment without loss of scale precision or content validity

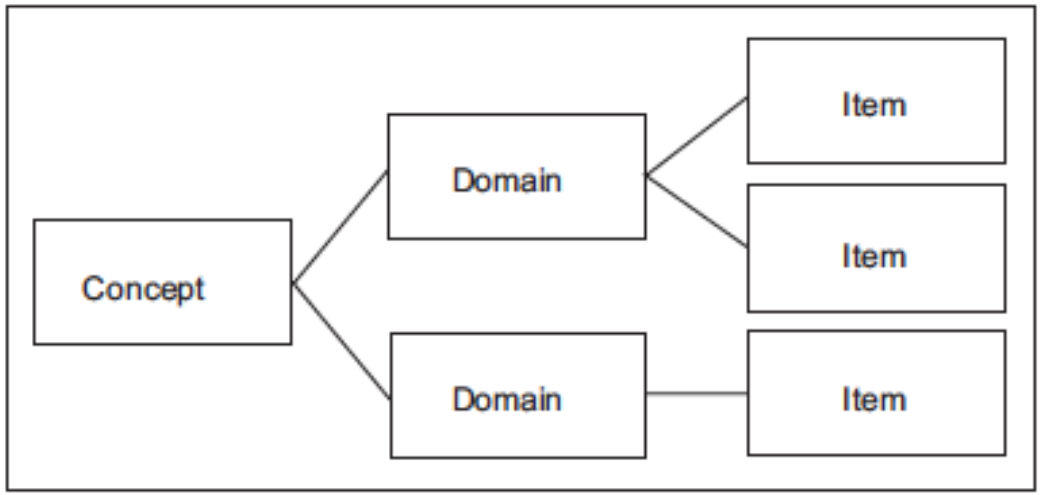


Figure 1: Conceptual framework in PRO instrument

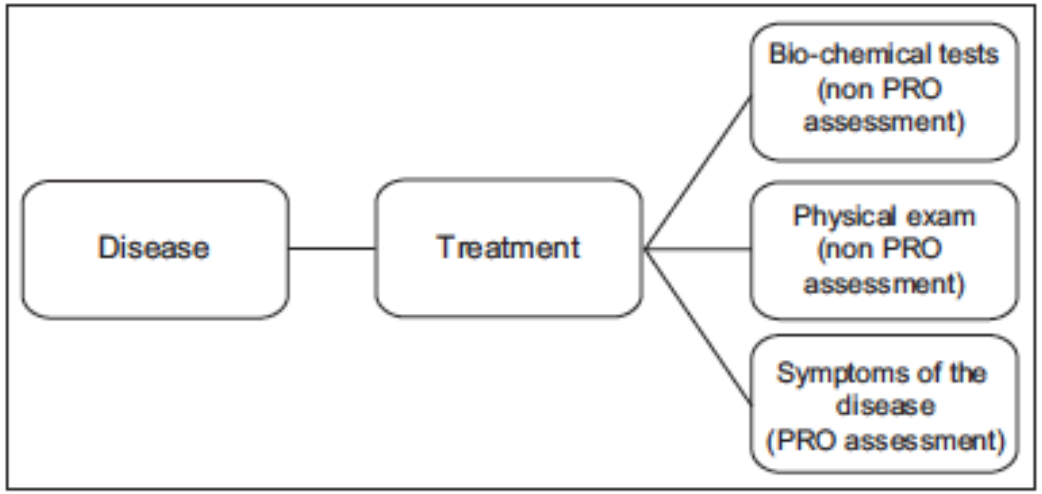


Figure 2: End-point model

FDA Guidance

- 2009 - FDA published guidance for industry for PROMs
  - Use in Medical Product Development to Support Labeling Claims
- Focus on what matters to the patient
  - Status of patient's health condition, directly from the patient
  - No interpretation of the response by clinician or anyone else
- Documentation of evidence of patient input during instrument **development**
- Documentation of patient input in **performance** of the instrument in specific application it will be used.

FDA guidance

- Specific guidance on development of instruments
  - Iterative process
  - Document iterative process
- Reasons for changing PROs (table from the FDA guidance)
- FDA review considerations

FDA Guidance

- 21CFR314.126(b)(6)

"The methods of assessment of subjects' response are well-defined and reliable. The protocol for the study and the report of results should explain the variables measured, the methods of observation, and the criteria used to assess response."

Item Property	Reason for Change or Deletion
Clarity or relevance	Reported as not relevant by a large segment of the target population Generates an unacceptably large amount of missing data points Generates many questions/requests for clarification from patients as they complete the PRO instrument Patients interpret items and responses in a way inconsistent with the instrument's conceptual framework
Response range	High percent of patients respond at the floor (scale's worst end) or ceiling (scale's optimal end) Patients note that none of the response choices applies to them Distribution of item responses is highly skewed
Variability	All patients give the same answer (i.e., no variance) Most patients choose only one response choice Differences among patients are not detected when important differences are known
Reproducibility	Unstable scores over time when there is no logical reason for variation from one assessment to the next
Inter-item correlation	Item highly correlated (redundant) with other items in the same concept of interest
Ability to detect change	Item is not sensitive (i.e., does not change when there is a known change in the concepts of interest)
Item discrimination	Item is highly correlated with measures of concepts other than the one it is intended to measure Item does not show variability in relation to some known population characteristics (i.e., severity level, classification of condition, or other known characteristic)
Redundancy	Item duplicates information collected with other items that have equal or better measurement properties
Recall period	Population, disease state, or application of the instrument can affect appropriateness of the recall period

FDA review considerations

- Reliability
- Validity
- Ability to detect change

FDA guidance

- Response option types
  - VAS, Likert, pictorial
- Respondent and administrator burden
  - Length of questionnaire/interview
  - Formatting/font size
  - New instructions for each item
  - Need to consult records to complete
  - Privacy in which the PRO completed
  - Inadequate time to complete
  - Literacy level to high
  - Questions patients unwilling to answer
  - Perception that interviewer wants/expects specific answer
  - Need for physical help to complete

FDA guidance

- Instrument modification
  - Change from paper to electronic
  - Changing timing of procedures for PRO admin
  - Changing to a different setting, population, or condition
  - Changing order of items, item wording, response options, recall period, or deleting portions of questionnaire
  - Changing instructions or placement of instructions in the PRO instrument

FDA guidance: clinical trial design

- Design and analysis:
  - Same for PROs as for any other endpoint
- General protocol considerations
- Frequency of assessments
- Clinical trial duration
- Design considerations for multiple endpoints
- Planning for clinical trial interpretation using a responder definition

- Specific concerns when using electronic PROs

What is MACRA?

- The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
- Passed with wide bipartisan and bicameral support
  - House vote: 392-37
  - Senate vote: 92-8
- Signed into law April 16, 2015

Merit-Based Incentive Payment System (MIPS)

Last reporting period = 2016
Last payments = 2018

EHR Incentive Program – attest per provider

"Meaningful Use"

PQRS and CAHPS – report as a group

(Physician Quality Reporting System + Clinician and Group – Consumer Assessment of Healthcare Providers and Systems

Value Modifier – report as a group

(measured with PQRS, CAHPS and claims data)

First reporting period = 2017
First payments = 2019

Merit-Based Incentive Payment System (MIPS) – report per provider

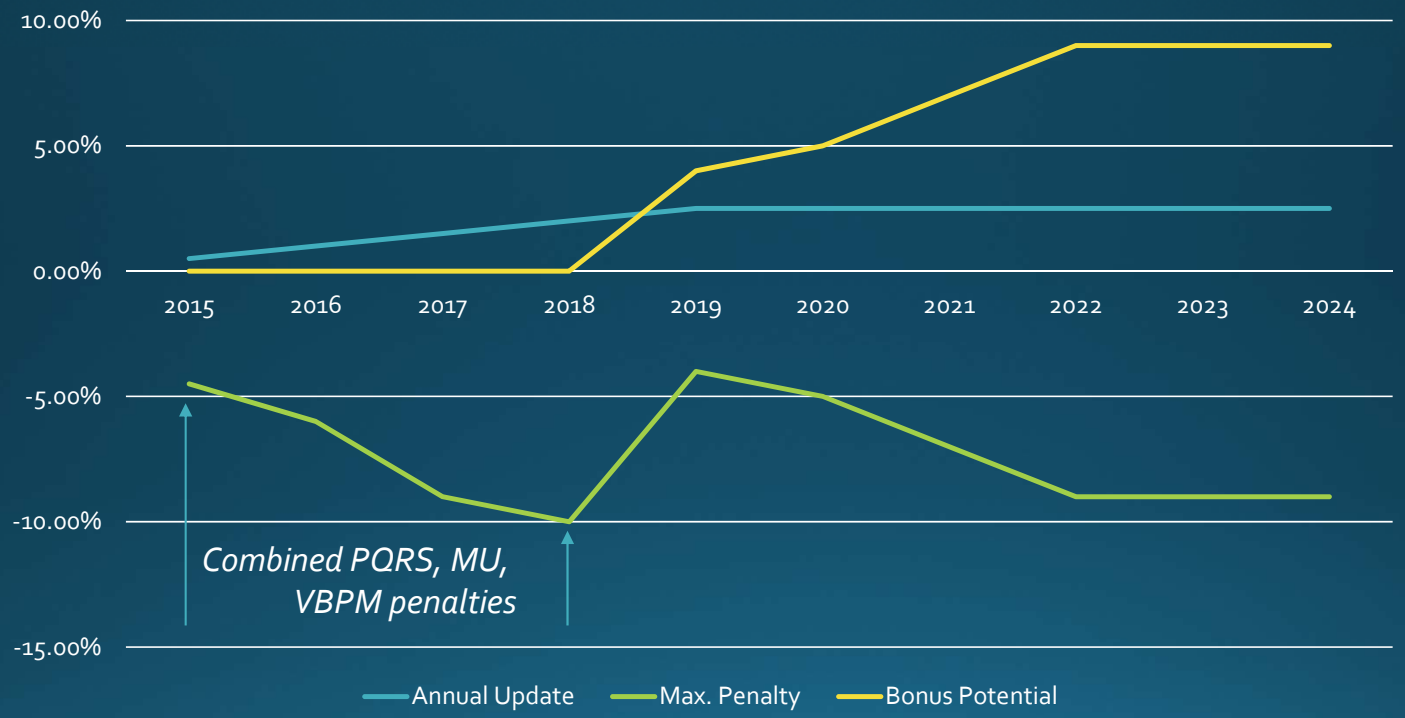
1. Meaningful Use
2. Quality
3. Resource Use
4. Performance Improvement

OR

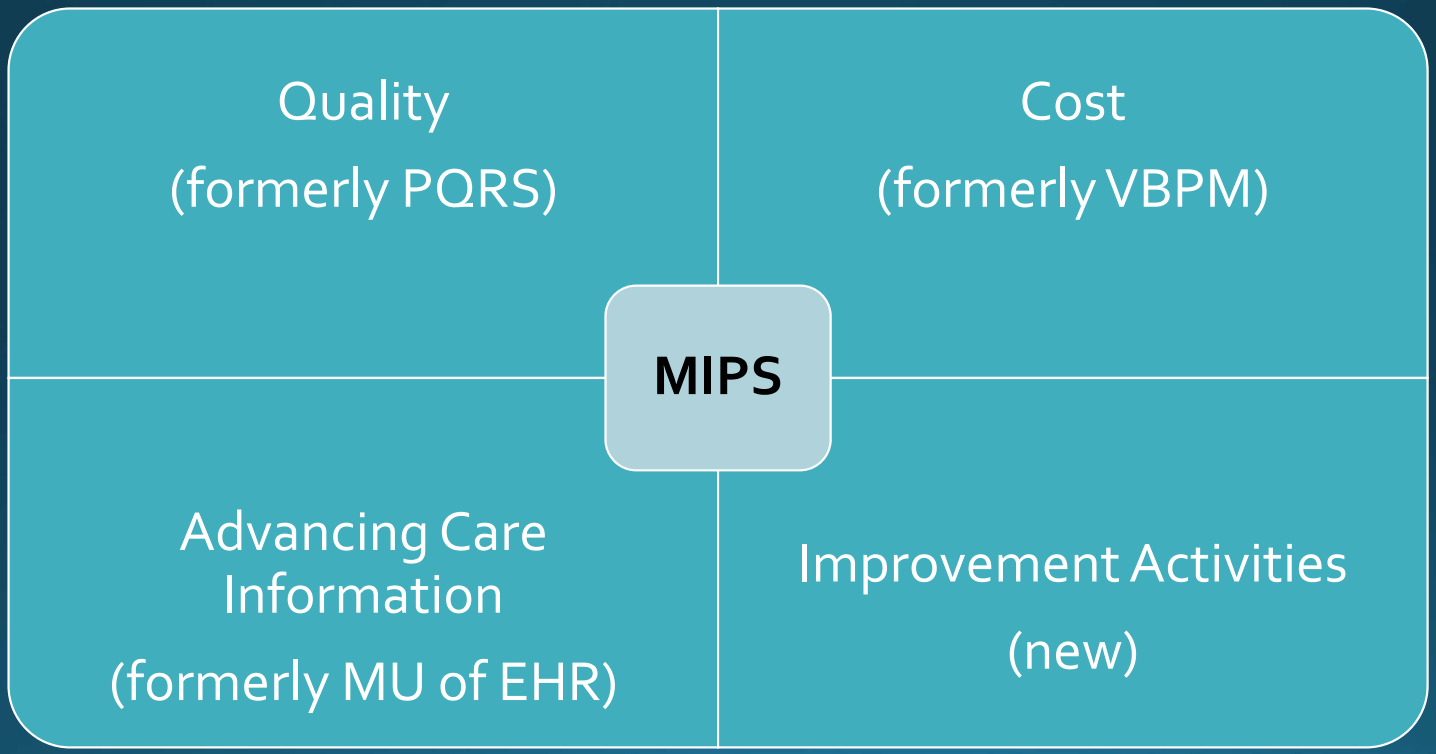
Participate in an **Alternate Payment Model (APM)** such as Accountable Care Organization, Comprehensive Primary Care, etc

Orange includes patient experience ratings

MIPS Risk Corridor



Components of MIPS



MIPS: Quality Category

- Report all 6 required measures
- Groups of 100+ can receive partial credit for electing to participate in the CAHPS for MIPS survey (1 cross-cutting and/or patient experience measure)

Search for measures to use on the CMS QPP site:

https://qpp.cms.gov/measures/performance

Explore Measures - Quality Me

https://qpp.cms.gov/measures/quality

Quality Payment Program

Learn About the Program | Explore Measures | Education & Tools

Program Performance | Quality Measures | Advancing Care Information | Improvement Activities

Quality Measures

Instructions

1. Review and select measures that best fit your practice.
2. Add up to six measures from the list below, including one outcome measure. You can use the search and filters to help find the measures that meet your needs or specialty.
3. If an outcome measure is not available that is applicable to your specialty or practice, choose another high priority measure.
4. Download a CSV file of the measures you have selected for your records.

Groups in APMs qualifying for special scoring standards under MIPS, such as Shared Savings Program Track 1 or the Oncology Care Model: Report quality measures through your APM. You do not need to do anything additional for the MIPS quality category.

Note: This tool is only for informational and estimation purposes. You can't use it to submit or attest to measures or activities.

2017 MIPS Performance

Select Measures

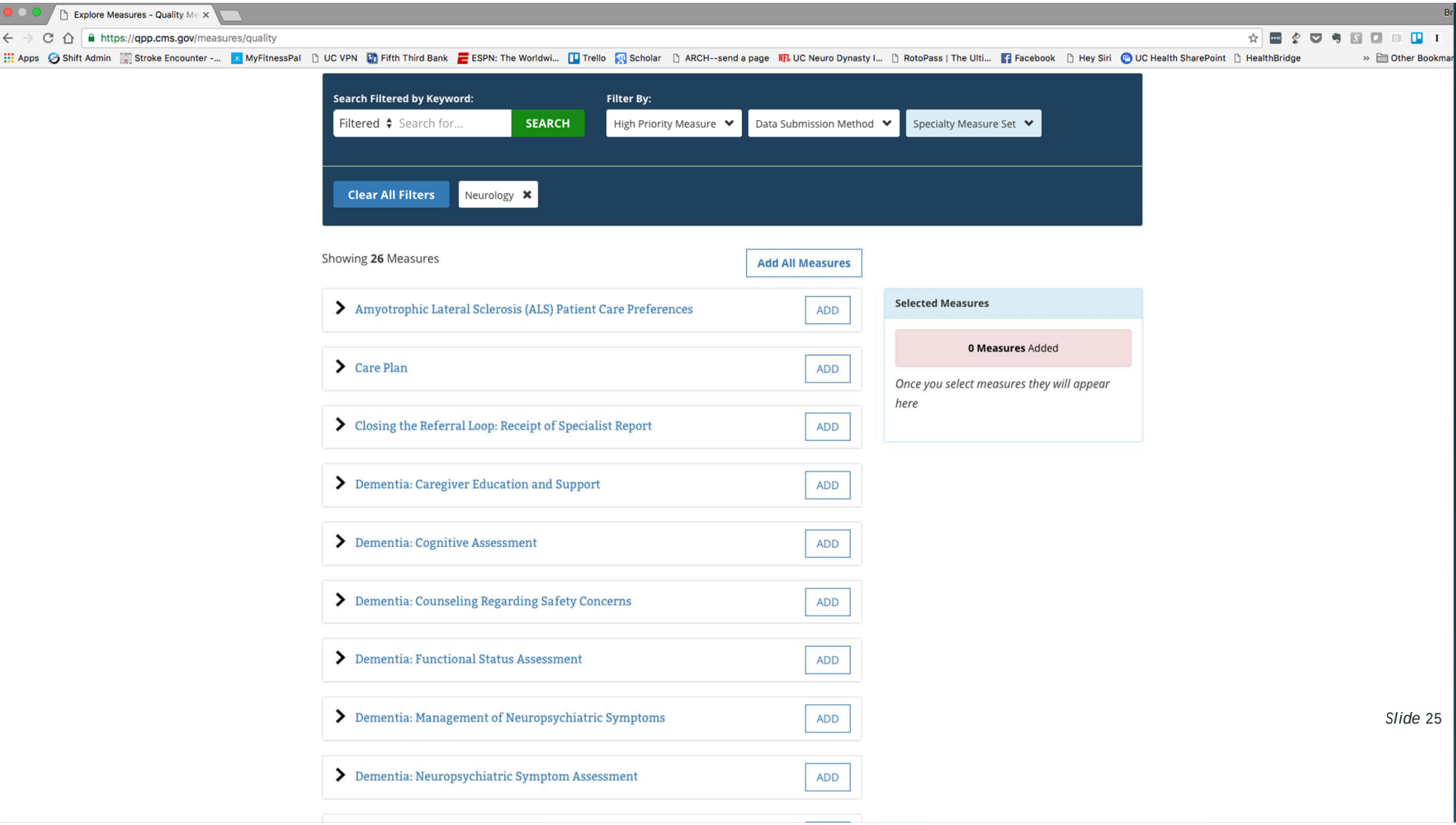
Search Filtered by Keyword:

Filtered | Search for... | SEARCH

Filter By:

High Priority Measure | Data Submission Method | Specialty Measure Set

Clear All Filters | Neurology ✖


Explore Measures - Quality Me
https://qpp.cms.gov/measures/quality
Apps
Shift Admin
Stroke Encounter -...
MyFitnessPal
UC VPN
Fifth Third Bank
ESPN: The Worldwi...
Trello
Scholar
ARCH--send a page
UC Neuro Dynasty I...
RotoPass | The Ulti...
Facebook
Hey Siri
UC Health SharePoint
HealthBridge
Other Bookmar
Search Filtered by Keyword:
Filtered
Search for...
SEARCH
Filter By:
High Priority Measure
Data Submission Method
Specialty Measure Set
Clear All Filters
Neurology
Showing 26 Measures
Add All Measures
Amyotrophic Lateral Sclerosis (ALS) Patient Care Preferences
ADD
Care Plan
ADD
Closing the Referral Loop: Receipt of Specialist Report
ADD
Dementia: Caregiver Education and Support
ADD
Dementia: Cognitive Assessment
ADD
Dementia: Counseling Regarding Safety Concerns
ADD
Dementia: Functional Status Assessment
ADD
Dementia: Management of Neuropsychiatric Symptoms
ADD
Dementia: Neuropsychiatric Symptom Assessment
ADD
Selected Measures
0 Measures Added
Once you select measures they will appear here

Quality of Life Assessment For Patients With Primary Headache Disorders

ADD

Percentage of patients with a diagnosis of primary headache disorder whose health related quality of life (HRQoL) was assessed with a tool(s) during at least two visits during the 12 month measurement period AND whose health related quality of life score stayed the same or improved

MEASURE NUMBER

- eMeasure ID: N/A
- eMeasure NQF: N/A
- NQF: N/A
- Quality ID: 435

NQS DOMAIN

Effective Clinical Care

MEASURE TYPE

Outcome

HIGH PRIORITY MEASURE

Yes

DATA SUBMISSION METHOD

- Claims
- Registry

SPECIALTY MEASURE SET

- Neurology

PRIMARY MEASURE STEWARD

American Academy of Neurology

Top

Development of a PROM

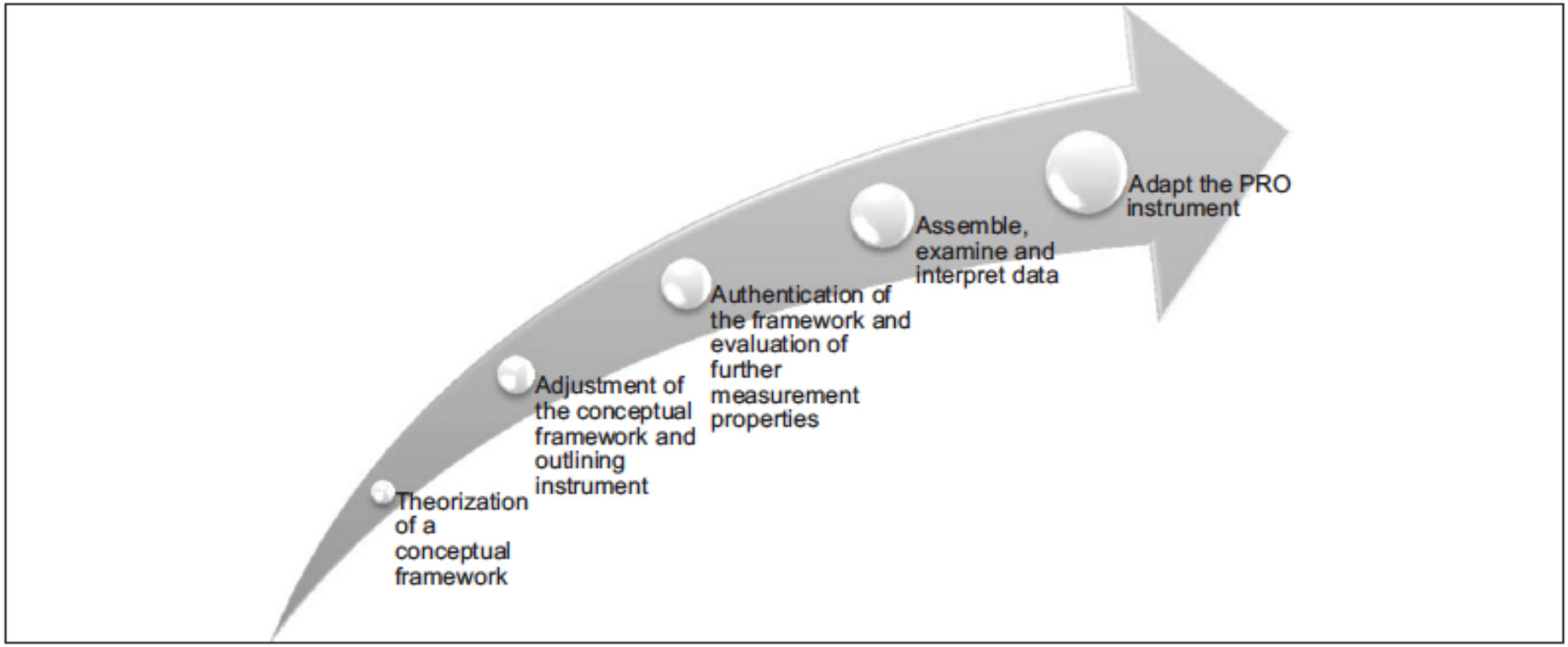
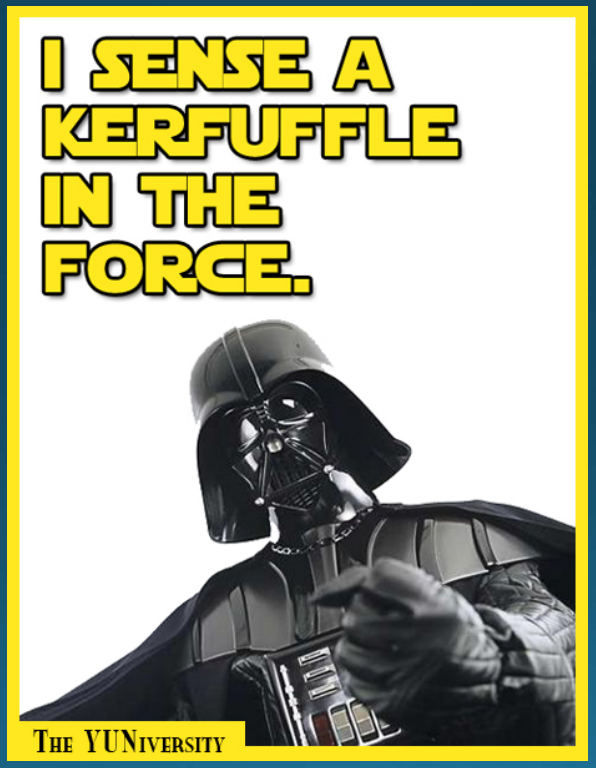


Figure 4: Development of PRO instrument

Examples from Real Life


I SENSE A KERFUFFLE IN THE FORCE.
THE YUNIVERSITY

PHQ-9

- Brett Kissela's clinic:
  - Many/most patients require depression screening
  - Paper form, patient fills out
  - Screen based on 2 major symptoms and total score for determining treatment
  - Can easily enter into EPIC, template exists
    - There is a dot phrase to enter it directly into note too

Promise of PROMIS/NeuroQOL

- K23: Outcome and Quality of Life after Diabetic Stroke

- Multiple measures: SF-36 (SF-12), EQ-5D, SSQOL
- Also CES-D, etc.

- Nurse coordinator mutiny

Spasticity Screening Tool

- Full disclosure, a project funded by Allergan
- Spasticity is Term for velocity-dependent increase in muscle tone accompanied by increased reflex activity
  - a vague term that providers variably recognize/diagnose and measuring is difficult
  - Modified at times to include "stiffness" but this may bleed into rigidity or dystonia
- Scales for assessing/measuring don't cross diseases well, are long and inconsistently measured
  - Ashworth, modified Ashworth, Penn Spasm Frequency, Modified Tardieu Scale, etc.

Spasticity Screening Tool

- None of the available measures was designed for screening
- Ideal = an easy tool for patient self-identification that prompts treatment as indicated
- Conceptual framework: content validity
  - Understandable terms for spasticity
  - Must apply to all limbs
  - Ideally applicable to multiple diseases
  - Domains of interest: pain, ADLs (hygiene, dressing), mobility

Methods

- Delphi panel
  - Identification of candidate items for screening tool from existing measures
  - Modified Delphi process to achieve maximal consensus
    - Survey to examine endorsement of items
    - Re-ranking after seeing anonymized results
    - Final meeting to achieve consensus--> drafting of final scale for cognitive debriefing/validation
    - 11 panel members, 47 items narrowed down to 13
    - Agreement that shorter is better (goal 11-15 items)—thus narrowed to 13 items that covered perceived signs/sx of spasticity and impact on function/QOL
- Patient Interviews

Methods

- Delphi panel
- Patient Interviews
  - 20 patients across 5 disease states
  - 10/13 items interpreted as intended by >90%
  - Other 3 items 80-85% correct interpretation
  - Slight modifications and examples added to several questions

  - FINAL SPASTICITY SCREENING TOOL
    - Zorowitz RD, Wein TH, Dunning K, Deltombe T, Olver JH, Davé SJ, Dimyan MA, Kelemen J, Pagan FL, Evans CJ, Gillard PJ, Kissela BM. A Screening Tool to Identify Spasticity in Need of Treatment. Am J Phys Med Rehabil. 2017 May;96(5):315-320

Next steps = validation

- Will assess:
  - Internal consistency with Cronbach's alpha
  - Test re-test reliability
  - Convergent validity
  - Classification accuracy of the tool relative to the gold standard diagnostic assessment in clinic (with receiver operating curves for identifying a cut score maximizing accuracy)
  - Ability to detect/measure change

CMT Neuropathy Score

- CMT slowly progressive neuropathy
- Slow change over time in clinical examination and in electrophysiology
- For clinical trials, development of meaningful outcome measures were of import
  - Neuropathy Impairment Score based on examination findings
  - Reproducibility good with training but not sensitive to change

CMT Neuropathy Score

- International Neuropathy Consortium: natural history data, find targets for treatment, initiate trials
- Need for outcome measures that could measure change over time
- Focus groups with patients to help develop patient reported outcomes
  - What was important to patient
  - What did they note changed over time
  - Development of questionnaires

CMT Neuropathy Score

- 2005 – validation of a tool: CMTNS
  - Series of CMT patients across the spectrum of disease and different types of CMT
  - Shy M et al, Reliability and validity of the CMTNS as a measure of disability. Neurology 64: 1209-14
- Ascorbic acid in CMT1A mouse model – improved demyelination
- Large multicenter trial with CMT1A in US, similar in UK/Italy
  - Both with negative outcomes
  - Historical controls with CMTNS did worse than placebo group in trial
  - Significant floor and ceiling effects

CMT Neuropathy Score

- International workshop to improve the score
  - Murphy SM et al. Reliability of the CMT neuropathy score (second version) in Charcot-Marie-Tooth disease. J Periph Nerv System 2011 (16):191-8
- Several small but significant changes
- PRO portion:
  - Script for interviewer to ensure question asked consistently
  - Sensory symptoms using a picture to standardized patient scoring
  - Sensory symptoms also ranged higher than above ankle
  - Motor symptoms: removed weight of ankle surgery, added weight for shoe inserts (decrease floor effect from orthopedic practices for early surgery)
  - Motor symptoms given a range
    - "mild difficulty with buttons" to "severe/unable to do buttons"
    - "unable to cut most food" from "unable to write/use keyboard"

CMT Neuropathy Score – Version 2

Parameter	0	1	2	3	4	Score
Sensory symptoms[1]	None	Symptoms below or at ankle bones	Symptoms up to the distal half of the calf	Symptoms up to the proximal half of the calf, including knee	Symptoms above knee (above the top of the patella)	
Motor symptoms legs[2]	None	Trips, catches toes, slaps feet. Shoe inserts	Ankle support or stabilization (AFOs). Ankle bone surgery or tendon transfers.	Walking aids (cane, walker)	Wheelchair	
Motor symptoms arms	None	Mild difficulty with buttons	Severe difficulty or unable to do buttons	Unable to cut most foods	Proximal weakness (affect movements involving the elbow and above)	
Pinprick sensibility[1,3]	Normal	Decreased below or at ankle bones	Decreased up to the distal half of the calf	Decreased up to the proximal half of the calf, including knee	Decreased above knee (above the top of the patella)	
Vibration[4]	Normal	Reduced at great toe	Reduced at ankle	Reduced at knee (tibial tuberosity)	Absent at knee and ankle	
Strength legs	Normal	4+,4 or 4- on foot dorsiflexion	≤ 3 on foot dorsiflexion	≤ 3 on dorsi and plantar flexion	Proximal weakness	
Strength arms	Normal	4+,4 or 4- on intrinsic hand muscles[5]	≤ 3 on intrinsic hand muscles[5]	< 5 on wrist extensors	Weak above elbow	
Ulnar CMAP (Median)	>6mV (>4mV)	4-5.9mV (2.8-3.9)	2-3.9 mV (1.2-2.7)	0.1-1.9 mV (0.1-1.1)	Absent (Absent)	
Radial SAP amplitude, antidromic	≥15µV	10 - 14.9 µV	5 - 9.9 µV	1 - 4.9 µV	< 1 µV	
Total						

Notes: 1: Use the picture below to discriminate the level of the symptoms; 2: Uses aid <u>most of the time.</u> The patient was prescribed to wear/use or should be wearing/using the aid in the examiner's opinion; 3: Abnormal if patient says it is definitely decreased compared to a normal reference point; 4: Use Rydell Seiffer tuning fork. Definition of Normal: ≥ 5; 5: Intrinsic hand muscles strength assessment: Test only Abductor Pollicis Brevis (ABP) and First Dorsal Interosseus (FDI), then choose the stronger to give the score.

Patient Name: _____ **Date:** _____ **Evaluator:** _____

CMT Neuropathy Score

- Comparison between CMTNS and CMTNS-2
  - Significant difference in mean scores for sensory and motor symptoms
  - Mild (0-11), moderate (12-21) and severe (22- 36)
  - Major categories the same
  - Mean scores for sensory symptoms lower, motor symptoms higher

  – more sensitive to minor differences/change over time

  Interestingly – more inter/intrarater discrepancy in the sensory examination than the PROs

Development of a PROM

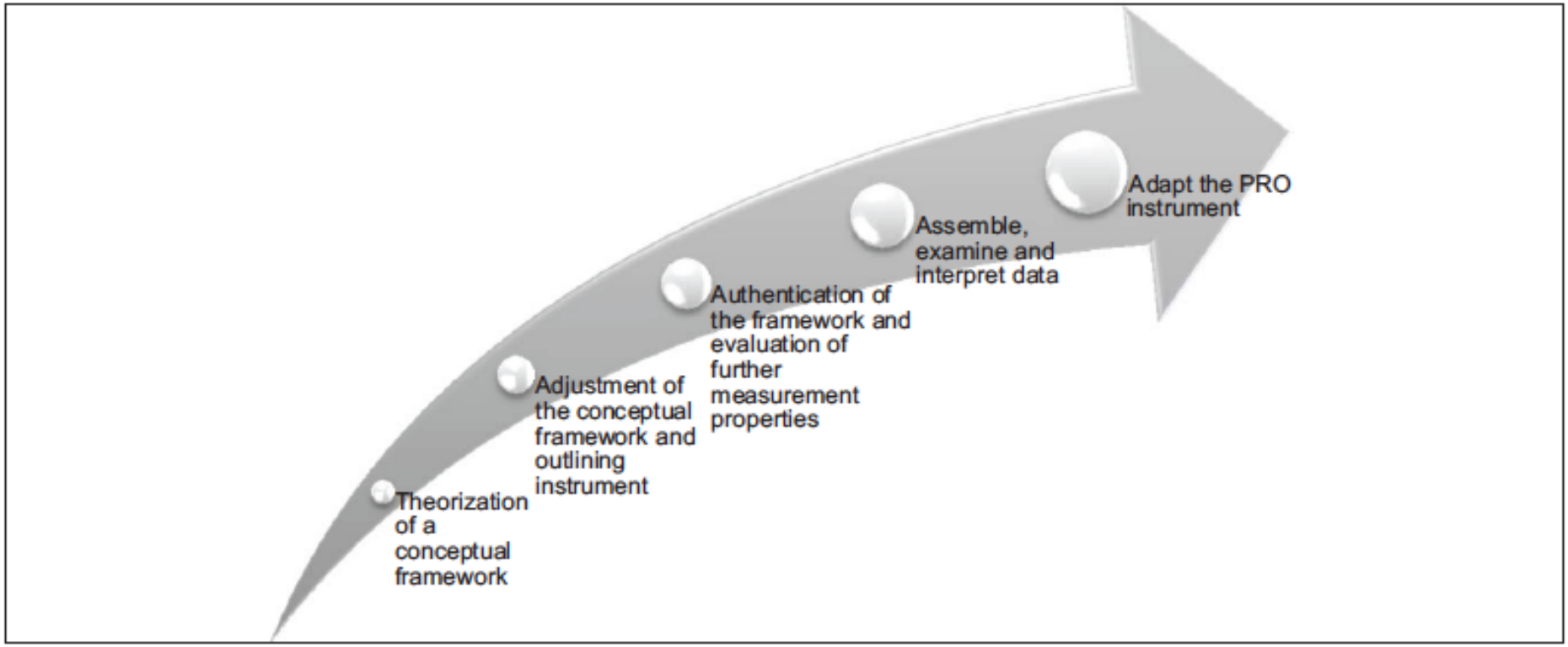


Figure 4: Development of PRO instrument

READING:

- PROMIS website, at a minimum the Overview page
- http://commonfund.nih.gov/promis/overview

- Check out Neuro-QOL for validated neuro outcome measures
  http://www.healthmeasures.net/explore-measurement-systems/neuro-qol