Patient-Reported Outcomes

Workshop

July 30-31, 2012

Project Scope and Activities

- Focused on methodological issues, not endorsement
- Under the guidance of an expert panel 2 workshops are planned to bring together the stakeholders necessary to facilitate the groundwork for the development, testing, endorsement, and implementation of PRO-based performance measures.
- Two commissioned papers will help inform next steps regarding: 1) selection of patient-level PROs for use in performance measures, and 2) the path to developing reliable and valid PRO-based performance measures eligible for NQF endorsement that can be used for accountability and to inform quality improvement
- Funded by HHS
### Timeline

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

### Quality Measurement Enterprise

- **Measure Development**
  - Priorities and Goals
  - Standardized Measures
  - Electronic Data Platform
  - Alignment of Environmental Drivers
  - Evaluation and Feedback

- **National Priorities Partnership**
  - High Impact Conditions

- **NQF Endorsement Process**

- **Quality Data Model**
  - eMeasures Format

- **Measures Applications Partnership**
  - Measures Database
  - Model Dashboard

- **NPP Evaluation**
  - Measure Use Evaluation
  - Measure Maintenance
Performance Measurement in Evolution

- Drive toward higher performance
- Measure disparities in all we do
- Shift toward composite measures
- Harmonize measures across sites and providers
- Measurement across longitudinal patient-focused episodes
  - Outcome measures (including PROs)
  - Process measures with direct evidence of impact on desired outcomes
  - Appropriateness measures
  - Cost/resource use measures coupled with quality measures, including overuse

Evidence for the Measure Focus

- Hierarchical preference for
  - Outcomes linked to evidence-based processes/structures
  - Outcomes of substantial importance with plausible process/structure relationships
  - Intermediate outcomes
  - Processes/structures
  
Most closely linked to outcomes
Individual-Level PRO vs. Performance Measure

- **NQF does not endorse individual-level instruments or scales**
  - Although reliable and valid and useful in clinical practice or research, individual patient scores alone are *not* sufficient to determine performance and make conclusions about quality of a healthcare entity
  - Individual-level scores are the data that would be used in a performance measure
- **NQF endorses performance measures that result in a score for an accountable healthcare entity (and use data from all eligible patients)**
  - An endorsed performance measure must be standardized and precisely specified so specific instruments/scales and scoring must be identified

Examples: Endorsed PRO Performance Measures

- **Depression** (MN Community Measurement)
  - Depression Utilization of the Patient Health Questionnaire (PHQ-9) tool paired with:
    - Depression remission at six months
    - Depression remission at twelve months
- **Visual Function** (AAO)
  - Improvement in patient’s visual function achieved within 90 days following cataract surgery
    - Improvement in visual function is defined by the quantitative scale used in the VF-14 survey instrument pre- and post-surgery.
Value Proposition

- Individual Level PRO
  - Inform care processes
  - Patient feedback and self monitoring
  - Shared decision-making
- Aggregate Level: Performance Measure
  - Quality improvement
  - Accountability (e.g., public reporting/transparency, payment)

Determinants of Health Model

Patient-Focused Episode of Care Model

PRO Categories Across the Episode:
- HRQOL/Functional Status
- Health-related Behaviors
- Symptom/Symptom Burden
- Patient Experience with health care
Today’s Meeting Objectives

- Identify best practices and lessons learned from initiatives that have implemented individual-level PROs in performance measurement;
- Discuss the major methodological issues related to the selection, administration and use of individual-level PROs in performance measures;
- Discuss key considerations for inclusion of PROs into EHRs and policy implications;
- Identify the characteristics of individual-level PROs suitable for potential use in performance measures; and
- Identify an initial set of PROs most suitable for development and testing of performance measures.
Patient-reported outcomes in health care performance measurement: Issues related to selection and administration

David Cella, PhD
Department of Medical Social Sciences
Feinberg School of Medicine
Northwestern University

Co-authors

- Elizabeth A. Hahn
- Sally E. Jensen
- Zeeshan Butt
- Cindy J. Nowinski
- Nan Rothrock
Why not just ask clinicians?

- Vast literature demonstrating that clinical providers do not accurately capture outcomes that are logically obtained by direct patient query
Comparison of Paired Observations

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Patient graded higher</th>
<th>Agreement</th>
<th>Clinician graded higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue</td>
<td>7%</td>
<td>34%</td>
<td>41%</td>
</tr>
<tr>
<td>Pain</td>
<td>7%</td>
<td>19%</td>
<td>60%</td>
</tr>
<tr>
<td>Constipation</td>
<td>3%</td>
<td>24%</td>
<td>69%</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>11%</td>
<td>85%</td>
<td>4%</td>
</tr>
<tr>
<td>Anorexia</td>
<td>10%</td>
<td>18%</td>
<td>66%</td>
</tr>
<tr>
<td>Nausea</td>
<td>3%</td>
<td>14%</td>
<td>77%</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>6%</td>
<td>26%</td>
<td>52%</td>
</tr>
<tr>
<td>Cough</td>
<td>4%</td>
<td>18%</td>
<td>67%</td>
</tr>
</tbody>
</table>

Clinician AEs and Patient Reports: Lung Cancer

- **Uniscale**: $r = -0.06$
- **Functional Assessment of Cancer Therapy Lung (FACT-L)**: $r = 0.10$
- **Lung Cancer Symptom Scale (LCSS)**: $r = -0.03$
- **Symptom Distress Scale (SDS)**: $r = -0.11$

Clinician-reported AEs and PROs measure different aspects of the disease/treatment experience and are complementary.
Potential for PRO use in clinical care

- Assist clinical providers in care management
- Enhance clinical efficiency
- Improve patient-provider communication
- Identify patient needs in a timely manner
- Facilitate patient-centered care

However...

- **Routine PRO assessment is not common in clinical practice**

Patient Experience of Care:
An integral component of patient-centered care

**Patient satisfaction**
(example from FACIT-TS; [www.facit.org](http://www.facit.org))

<table>
<thead>
<tr>
<th>Did your doctor seem to understand what was important to you?</th>
<th>No, not at all</th>
<th>Yes, but not as much as I wanted</th>
<th>Yes, almost as much as I wanted</th>
<th>Yes, and as much as I wanted</th>
</tr>
</thead>
</table>

**Patient reports of their actual experiences**
(example from CAHPS; [www.ahrq.gov/cahps](http://www.ahrq.gov/cahps))

In the last 12 months, when you phoned this provider’s office during regular office hours, how often did you get an answer to your medical question that same day?

<table>
<thead>
<tr>
<th>Never</th>
<th>Sometimes</th>
<th>Usually</th>
<th>Always</th>
</tr>
</thead>
</table>
Figure 1.1 Based upon Epstein et al, *Soc Sci Med*, 2005, [from 2007 NCI/NIH Pub. #07-6225 “Patient-Centered Communication in Cancer Care”]

Best practices to minimize self-report barriers

- Select appropriate method and mode of administration
  - Consider age, functional status, cognition as they relate to use of proxies and assistive devices.

- Universal design principles, quality translations and cultural adaptations, provide equivalent versions

- Flexibility in location (in-clinic, at home, at facilities)
  - requires access to the technology selected for the PRO
  - health information privacy must be protected

- Addressing *functional literacy* and *health literacy* are critical to delivering person-centered health care
Best practices to minimize barriers to self-reporting

- There are some circumstances in which it may be difficult or impossible to directly obtain PRO assessment by self-report.

- Proxy reporting can be useful:
  - for people with cognitive or communications deficits or severe disease burden
  - for people in the early stages of dementia who may fail to recognize the extent of their impairment
  - for young children
Methodological Issues: Method of Administration, Collection & Response
Implications of Method/Mode of Administration and Response

- Decisions must be made related to data collection methods and the implications of those decisions on costs and errors in surveys (Groves, 2009)
  - What is the most appropriate method to choose for a particular question?
  - What is the impact of a particular method on survey errors and costs?
- Methods and modes differ along various dimensions:
  - Degree of interviewer involvement
  - Level of interaction with respondent
  - Channels of communication used
  - Degree of technology use

Source of Data and Methods/Modes of Survey Administration

- Self-report vs. proxy/observer

Self-administration
  - paper-and-pencil
  - telephone
  - computer

Interviewer-administration
  - paper-and-pencil
  - telephone
  - computer
Implications of Data Source: Self Versus Proxy

- Proxy reporting is useful when difficult or impossible to obtain PROs directly
  - Allows broader inclusion and more representative range of patients
  - Minimizes missing data and increases the feasibility of longitudinal assessment

- Proxy reports may substitute for or complement patient assessment
  - May involve proxies assessing the patient as they think the patient would respond
  - May involve proxies providing their own perspective on the patient’s status

- Evaluating agreement between patients and proxies
  - Greater agreement when rating observable functioning, activities of daily living, physical health, motor functioning and less agreement when rating social functioning, pain, cognitive status, psychological, emotional well-being
  - Magnitude of disagreement can be minimized
    - Disagreement may provide useful information (e.g., MCI → early dementia)

Implications of Mode of Administration

- Mode choices involve trade-offs and compromises
  - Consider the objectives of the assessment and the resources available
    - Self administration:
      - Advantages: Cost; May yield more participant disclosure; Proceed at one’s own pace
      - Disadvantages: Potential for missing data, Requires simple survey design
    - Interviewer administration
      - Advantages: Allows more complex survey design, Useful for patients with reading, writing, or vision difficulties
      - Disadvantages:: Cost; Potential for bias

- Concern about the effects of mode on data quality
  - Reliability is high for both
  - Response effects favor self-administration but inconsistent
Implications of Method of Administration

- **Paper-and-pencil**
  - Advantages: Low start-up cost
  - Disadvantages: prone to data entry errors, data entry and scoring require more time, hard to incorporate into EHR

- **Electronic**:  
  - Advantages: interactive, practical, increased comfort for socially undesirable behaviors, minimizes data entry errors, immediate scoring/feedback, easy to incorporate into EHR
  - Disadvantages: Up-front cost, potential discomfort with technology and accessibility

- **Potential for differences between p & p versus electronic capture:**
  - Impersonality of the method
  - Cognitive burden on patient
  - Control over the questionnaires
  - Communication style

- **Increasing evidence of measurement equivalence between methods**
  - As new methods are developed, it is critical to compare them to existing methods
  - Across methods, patient privacy is always a concern

PROMIS Example: Method of Administration
No meaningful differences found between modes of administration

< 1.5 points on 100-point scale

Comparisons to PC Administration: MID > 2 points (0.2 SD)
People preferred the computer screen interface

Implications of Setting of Administration

- **Clinic/Provider setting:**
  - **Strengths:**
    - Real-time assessment
    - Easy to implement electronic administration
  - **Limitations:**
    - Impact on clinic flow
    - Interruptions result in missing data
    - Patient distraction/anxiety
    - Staff burden

- **Home setting:**
  - **Strengths:**
    - Minimizes impact on clinic flow
    - Minimizes staff burden
  - **Limitations:**
    - Accessibility
    - Health information privacy
    - Data security
    - Patient safety
Non-Response and Response Shift

- Bias may be introduced by missing data
  - Evaluate the amount, reasons and patterns of missing data
  - Apply statistical adjustments based on degree and pattern of missing data
  - Strategies to evaluate non-response bias:
    - Conduct an abbreviated follow-up survey with initial non-responders
    - Compare characteristics of respondents and non-respondents
    - Compare respondent data with comparable information from other sources
    - Compare early versus late respondents

- Patient adaptation and response shift over time can complicate the interpretation of PRO outcomes
  - Improvement may be unrelated to treatment effect
  - Consider monitoring for response shift or implementing control/comparator arms with longitudinal follow-up

Implications of Test Theory Type

- Classical Test Theory (CTT): estimates the level of an attribute as the sum of responses to individual items
  - “Test-dependent”: validity dependent upon all items to be completed

- Item Response Theory (IRT): “test-free”: enables estimation of the latent trait using different items as long as their locations have been calibrated on the same scale as the patients’ ability levels

- IRT enables customized assessment, including computer-adaptive testing (CAT) in which the questions are tailored to the individual patient
  - Questionnaires can be shorter
  - The scale scores can be estimated more precisely for any given test length
  - Patients do not need to complete the same set of items in every situation
An item bank is a large collection of items measuring a single domain.

Any and all items can be used to provide a score for that domain.

The PROMIS Metric

T Score

Mean = 50
SD = 10

Referenced to the US General Population
PROMIS Basic Tools

Derived from Item Banks

Computerized Adaptive Testing (CAT)
- Dynamic testing averaging 6 items per domain

Fixed Length Forms
- By individual domain (8-10 items)
- By health profile (-29, -43, -57)

Global Health Index

Beginning of CAT

T-Score = 50  SE = 10

Best Item - I felt depressed

T-Score = 50  SE = 10
I felt depressed
1. Never
2. Rarely
3. Sometimes
4. Often
5. Always

T-Score = 52  SE = 4

I felt like a failure
1. Never
2. Rarely
3. Sometimes
4. Often
5. Always

T-Score = 53  SE = 3
I felt worthless
1. Never
2. Rarely
3. Sometimes
4. Often
5. Always

T-Score = 55  SE = 2

I felt unhappy
1. Never
2. Rarely
3. Sometimes
4. Often
5. Always

T-Score = 54  SE = 2
### Fatigue Item Bank

- **Cancer chemotherapy**
  - *Items 1-10*
  - Same metric, same meaning

- **Osteoarthritis program**
  - *Items 6-12*
  - Cat

- **Heart failure program**
  - *Items 1-5*

- **Pain management**
  - *Items 2, 4, 9, 13*

### PROMIS Measures Tested in Six Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Relevant Item Banks</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPD</td>
<td>Physical Function, Fatigue, Pain, Social Role Satisfaction, Emotional Distress (Depression, Anxiety, Anger)</td>
</tr>
<tr>
<td>Heart Failure</td>
<td>Physical Function, Fatigue, Social Role Satisfaction, Depression</td>
</tr>
<tr>
<td>Low Back Pain</td>
<td>Pain (Interference and Behavior), Physical Function, Depression, Fatigue, Sleep Disturbance</td>
</tr>
<tr>
<td>Depression</td>
<td>Emotional Distress (Depression, Anxiety, Anger), Sleep Disturbance, Fatigue, Physical Function</td>
</tr>
<tr>
<td>Cancer</td>
<td>Pain, Fatigue, Emotional Distress (Depression, Anxiety), Physical Function</td>
</tr>
</tbody>
</table>
PROMIS Fatigue Across Five Clinical Conditions

- **N = 310**: Cancer w/ benefit (2 mo) to Cancer Chemo (B)
- **N = 229**: Back Pain (3 mo) to Back Pain (B)
- **N = 114**: Depression (3 mo) to Depression (1 mo) (B)
- **N = 64**: HF Post-transplant to HF Pre-transplant
- **N = 125**: Exacerbation to Stable to COPD Stable (B) to COPD Exacerbation (B)

Average for General Population

PRO Rosetta Stone
PROsetta Stone Early Output

<table>
<thead>
<tr>
<th>FACIT-F Score</th>
<th>PROMIS T-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>27.8</td>
</tr>
<tr>
<td>51</td>
<td>32.8</td>
</tr>
<tr>
<td>50</td>
<td>35.9</td>
</tr>
<tr>
<td>49</td>
<td>38.4</td>
</tr>
<tr>
<td>48</td>
<td>40.3</td>
</tr>
<tr>
<td>47</td>
<td>42.0</td>
</tr>
<tr>
<td>46</td>
<td>43.4</td>
</tr>
<tr>
<td>45</td>
<td>44.8</td>
</tr>
<tr>
<td>44</td>
<td>45.8</td>
</tr>
<tr>
<td>43</td>
<td>46.9</td>
</tr>
<tr>
<td>42</td>
<td>47.9</td>
</tr>
<tr>
<td>41</td>
<td>48.8</td>
</tr>
<tr>
<td>40</td>
<td>49.8</td>
</tr>
<tr>
<td>39</td>
<td>50.5</td>
</tr>
<tr>
<td>38</td>
<td>51.3</td>
</tr>
<tr>
<td>37</td>
<td>52.1</td>
</tr>
<tr>
<td>36</td>
<td>52.8</td>
</tr>
<tr>
<td>35</td>
<td>53.6</td>
</tr>
<tr>
<td>34</td>
<td>54.3</td>
</tr>
<tr>
<td>33</td>
<td>55.0</td>
</tr>
<tr>
<td>32</td>
<td>55.7</td>
</tr>
<tr>
<td>31</td>
<td>56.3</td>
</tr>
<tr>
<td>30</td>
<td>57.0</td>
</tr>
<tr>
<td>29</td>
<td>57.6</td>
</tr>
<tr>
<td>28</td>
<td>58.3</td>
</tr>
<tr>
<td>27</td>
<td>58.9</td>
</tr>
<tr>
<td>26</td>
<td>59.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACIT-F Score</th>
<th>PROMIS T-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>60.2</td>
</tr>
<tr>
<td>24</td>
<td>60.8</td>
</tr>
<tr>
<td>23</td>
<td>61.4</td>
</tr>
<tr>
<td>22</td>
<td>62.1</td>
</tr>
<tr>
<td>21</td>
<td>62.7</td>
</tr>
<tr>
<td>20</td>
<td>63.4</td>
</tr>
<tr>
<td>19</td>
<td>64.0</td>
</tr>
<tr>
<td>18</td>
<td>64.6</td>
</tr>
<tr>
<td>17</td>
<td>65.3</td>
</tr>
<tr>
<td>16</td>
<td>65.9</td>
</tr>
<tr>
<td>15</td>
<td>66.6</td>
</tr>
<tr>
<td>14</td>
<td>67.3</td>
</tr>
<tr>
<td>13</td>
<td>68.0</td>
</tr>
<tr>
<td>12</td>
<td>68.8</td>
</tr>
<tr>
<td>11</td>
<td>69.5</td>
</tr>
<tr>
<td>10</td>
<td>70.4</td>
</tr>
<tr>
<td>9</td>
<td>71.2</td>
</tr>
<tr>
<td>8</td>
<td>72.1</td>
</tr>
<tr>
<td>7</td>
<td>73.0</td>
</tr>
<tr>
<td>6</td>
<td>74.1</td>
</tr>
<tr>
<td>5</td>
<td>75.3</td>
</tr>
<tr>
<td>4</td>
<td>76.5</td>
</tr>
<tr>
<td>3</td>
<td>77.9</td>
</tr>
<tr>
<td>2</td>
<td>79.7</td>
</tr>
<tr>
<td>1</td>
<td>81.9</td>
</tr>
<tr>
<td>0</td>
<td>85.0</td>
</tr>
</tbody>
</table>


Discussion
Methodological Issues: Selecting Patient-level PROs

Selecting PROs for Performance Measurement

- To optimize decision-making in clinical care, PROs must be measured in a standardized way using questionnaires with known properties
  - Many guidance documents address attributes for PROs used in research
  - Little guidance regarding attributes for PROs used as performance measures
Selecting PROs for Performance Measurement

- Differences in selecting PROs as performance measures vs research
  - More similarities than differences
  - Importance of shorter instrument length
  - Higher stakes (consequences)

- Established PROs have more evidence than newer PROs ...but newer PROs have better measurement properties
  - SF-36, SF-12, VR-36, VR-12 have been used most often
    - Limitation: Static measures
  - Future direction: IRT-based measures (e.g., PROMIS)

Characteristics of PROs Suitable for Use in Performance Measurement

- Review of recommended characteristics for PROs for use in performance measures

  - Example: Apply recommended PRO characteristics to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC; Bellamy, 2008)
    - PRO for use in individuals with knee and hip osteoarthritis
    - 24 items covering 1-14 days
    - 5-point Likert-type and 100mm visual analog formats available
    - 3 subscales:
      - Pain (5 items)
      - Disability/Physical Function (17 items)
      - Joint stiffness (2 items)
Characteristics of PROs Suitable for Use in Performance Measurement

1. Conceptual and measurement model

- Documentation should include description of:
  - Concept(s) included and the intended population(s)
  - Organization of concept(s) into a measurement model

- Target PRO should be a high priority for the health care system

- Example: WOMAC
  - Factorial validity of the physical function and pain subscales has been inadequate (Pua et al., 2009)

2. Reliability

- Internal consistency reliability should be:
  - ≥ 0.70 for group-level purposes
  - ≥ 0.90 for individual-level purposes

- Stability/Reproducibility depends upon the time window

- Example: WOMAC
  - Cronbach’s alpha for the three subscales range from 0.86 to 0.98
  - Stability has been adequate for the pain and physical function subscales, but less adequate for the stiffness subscale
3. **Validity**

- Evidence supporting:
  - Content validity
  - Construct validity
  - Criterion validity

- Limited number of PRO instruments have been validated for use in performance measurement

- **Example: WOMAC**
  - Development involved expert clinician input, and survey input from patients, as well as a review of existing measures
  - Patient ratings of satisfaction with arthroplasty correlate positively with WOMAC scores

4. **Responsiveness**

- Evidence of changes in scores consistent with predefined hypotheses regarding changes in the target population

- Important for performance measurement because there is an expectation of consequences

- Responsiveness is necessary if results are to be actionable

- **Example: WOMAC**
  - Demonstrates adequate responsiveness and ability to detect change in response to clinical intervention
Characteristics of PROs Suitable for Use in Performance Measurement

5. Interpretability of scores
   ▫ Documentation should include:
     ➢ What low and high scores represent
     ➢ Representative mean and SD in the reference population
     ➢ Guidance on estimating meaningful differences and change over time
   ▫ Performance measures:
     ➢ If different PROs are used, important to establish a link or cross-walk
     ➢ Application of criteria to determine clinically meaningful change
   ▫ Example: WOMAC
     ➢ Availability of population-based, age- and gender-normative values
     ➢ Availability of minimal clinically important improvement values
     ➢ Can be translated into utilities for economic evaluations

6. Burden
   ▫ Time, effort, and other demands on the respondent and the administrator
   ▫ Performance Measures:
     ➢ PRO assessments should be as brief as possible
     ➢ Reporting should be done in real-time
   ▫ Example: WOMAC
     ➢ Short form available
     ➢ Average time to complete mobile phone WOMAC = 4.8 minutes
Characteristics of PROs Suitable for Use in Performance Measurement

7. Alternative modes/methods of administration

- The use of multiple modes and methods can be useful for diverse populations
- Performance measures: Evidence of measurement equivalence necessary

- Example: WOMAC
  - Validated mobile phone and touchscreen platforms

Characteristics of PROs Suitable for Use in Performance Measurement

8. Cultural and language adaptations

- Performance measures: Mode, method and question wording must yield equivalent estimates of PRO measures

- Example: WOMAC
  - Available in over 65 languages
9. Electronic health records

- Performance measures: Critical features:
  - Interoperability
  - Automated, real-time measurement and reporting
  - Sophisticated analytic capabilities

- **Example:** WOMAC
  - Electronic data capture may allow for integration within EHR

---

Discussion
Incorporating PROs into Electronic Health Records & Personal Health Records

E-health:
“Health-related Internet applications that deliver a range of content, connectivity and clinical care”

- health information
- online formularies and prescription refills
- appointment scheduling and test results
- advance care planning and health care proxy designation

➢ e-health applications tend to focus on the needs of health care providers and organizations
➢ there is little evidence regarding whether the services offered are those that patients desire
Integrating PROs into EHRs & PHRs

- PROs will likely constitute an important aspect of future stages of “meaningful use” of EHRs

- Critical features:
  - Interoperability and widespread health information exchange
  - Automated, real-time quality and cost measurement
  - Sophisticated analytic capacities

- Important issues:
  - Patient perspective:
    - Patients want to be involved as a partner in the flow of information
  - Clinical buy-in
  - Compatibility with clinical flow
  - Meaningful use
  - Patient privacy:
    - Physical transfer of the paper-based PRO measure from patient to provider
    - Electronic transfer of data or unauthorized access to patient-reported data

- Key design principles:
  - Fitting PRO measures into flow of care
  - Designing the system with stakeholder engagement
  - Merging PRO data with other types of data
  - Engaging in continuous improvement of the systems based on user experience and new technology
Discussion
Physical Functioning (T-Score; Mean=50, SD=10)

Reliability/Precision of PROMIS (Physical Function)

Relative Precision of PROs (Physical Function)
Lessons from the field
Early experience with PROs

Partners HealthCare, Inc.
Boston, MA

Elizabeth Mort, MD, MPH
Senior Medical Director PHS
VP Quality & Safety MGH

Partners Strategic Plan: Care Redesign

- October 2010 launched Strategic Plan at Partners
- Care Redesign
  - Primary care, population health
  - Condition specific care
    - CABG
    - Stroke
    - Colectomy for Colon Cancer
    - AMI
    - Diabetes
### Key Guiding Principle behind Care Redesign

- **Outcomes**
  - Defined by patient
  - Measured for patient’s condition over entire episode of care

- **Value for Patients**
  - Health Outcomes
  - Cost of delivering outcomes

- **Cost**
  - Measured for patient’s condition over entire episode of care

---

### First wave of conditions: CABG & AVR and Diabetes

**Goals:** electronic, validated instruments, short, align

<table>
<thead>
<tr>
<th>CABG</th>
<th>Pre-Procedure</th>
<th>Post-Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FUNCTIONAL STATUS (PROMIS-10)</strong>&lt;br&gt;(General, mental, social, physical, anxiety, fatigue, pain)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>SYMPTOMS LEVEL</strong>&lt;br&gt;(chest pain, shortness of breath)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><strong>PERCEIVED HEALTH BENEFITS</strong>&lt;br&gt;(Perception of procedures’ success, and physical/emotional improvement compared to yr ago)</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td><strong>HEALTH UTILITY</strong>&lt;br&gt;(Health state from 1-100)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Total Questions:** 17<br>**Diabetes: Total Questions:** 21

<table>
<thead>
<tr>
<th>Diabetes</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FUNCTIONAL STATUS (PROMIS-10)</strong>&lt;br&gt;(General, mental, social, physical, anxiety, fatigue, pain)</td>
<td>10</td>
</tr>
<tr>
<td><strong>ANXIETY (PROMIS)</strong>&lt;br&gt;(worries, ability to focus, fearfulness)</td>
<td>4</td>
</tr>
<tr>
<td><strong>BURDEN OF DIABETES</strong>&lt;br&gt;(quality of life in light of disease, ability to cope w/ disease)</td>
<td>1</td>
</tr>
<tr>
<td><strong>HEALTH UTILITY</strong>&lt;br&gt;(Health state from 1-100)</td>
<td>1</td>
</tr>
</tbody>
</table>

**CABG: Pre-op Post-op 3,6,12 months**

**Diabetes: baseline and every six months**
Collaborated with a vendor for data collection

We have been in the field since March 2012

- **264** questionnaires completed since March 2012 (122 for cardiac surgery, CABG and AVR, and 142 for Diabetes).

- ~**56%** of the patients choosing a method of follow up selected the patient portal/internet option (not Interactive Voice Response).
Developing reporting formats and mechanisms, using electronic medical record and patient portal

The patient experience:
- Patients say their doctors “should be asking these questions”
- Patients comment that the tablets are fun to use and “very user-friendly”
- Patients are willing to answer these questions at home

The staff experience:
- Practice Administrators have created unique workflow plans for their clinics to best incorporate PROMs
- MA’s and nurses generally understand the importance and provide guidance to the patients through the process

The physician experience (preliminary):
- Variable response – from champions to critics
- Concern about fitting these data elements into the clinical encounter
- Concern about critical results and timely intervention
Lessons learned from planning and evaluation

- Integrate this data collection with system-wide focus on improving value
- Significant up-front investment in research, interviews, patient focus groups, change management
- Establish senior executive, system wide clinical champions, local physician leadership, operational champions
- Careful attention to each clinic’s unique workflow and organizational culture
- Engage providers in the design of reporting tools
Promise of PROs in Improving Patient Outcomes: Lessons from the Dartmouth Spine Center

Eugene C. Nelson, DSc, MPH
The Dartmouth Institute
Dartmouth-Hitchcock Health


Feed Forward PROs to Improve Outcomes and Health Care Value: Dartmouth Spine Center Case

Q: How is a kilowatt hour of electricity like a day in the hospital?
A: Nobody wants either. We want
- Cold beer & hot showers
- Better outcomes, better care & lower costs

- End use, least cost
- Value for money
  - Amory Lovins

Skating to where the puck is going to be … person-centered high value care
What is health care value?

**Population of Patients**

**A Health System**

**Value**

<table>
<thead>
<tr>
<th>Entry</th>
<th>Asmt</th>
<th>Dx</th>
<th>Rx</th>
</tr>
</thead>
</table>

**Initial Health Status**

**Healthcare Delivery**

**New Health Status + $$**

Value = Health outcomes (disease + risk + function) / costs over time

---

**Dartmouth Spine Center**

- Started in 1998 by Jim Weinstein
- Innovative interdisciplinary clinical microsystem ... 1 stop shopping
- “Back to work back to play 1 back at a time.” ... patient-centered
- Better care in real time & better research over time
- “I can’t be a good doctor if I don’t have PROMs”

---

Dr. Jim Weinstein, orthopod & D-H President
Spine Center: Feed forward (& feedback) system, featuring PROs for engaging patient, shared decision making & making care plan, coordinating care, improving care, measuring, researching & paying for health care value

Feed Forward

Referral or Visit Request
Orientation & PROs
Initial Work Up
Plan of Care

Feed Back

Improvement registry
Public reports website
SPORT & research

The summary report generated from patient-reported data is critical to a physician's ability to care for a patient

Patient Experience: "My" Outcomes

Risk Status
Disease Status
Functional Status

Herniated Disk Outcomes @ 2 Years

Surgery
41 Ave Age
43% Female

Non-Surgery
44 Ave Age
45% Female

Functional Costs

Physical SF-36 Improvement

Cost Per Quality Adjusted Life Year Added
By Surgery $74,870

Reduced Oswestry Symptoms

Functional Satisfaction

Total Direct & Indirect Costs

Going from Concept of value To measured value

Prototype SPORT Calculator

Degenerative Spondylolisthesis Treatment Calculator

Personalized risk assessment
Based on people like me ...
From research back to patient care
Patient History

Scores over time

SUCCESSES
Early Adopters

18 patient populations

Summary: to improve outcomes & value must capture PROs in flow of care and use it to improve outcomes and value of care for individual patients and populations

Lessons

- Patients’ reaction: 84% positive*
  - “Visit became very helpful, thorough & informative”
- Providers’ reactions*
  - “Patients get more involved in their care.”
  - “This changes how care is delivered.”
- Sustainable & replicable
  - 10 years at Spine Center & 18 DH programs & 70,000 patients
  - 13 SPORT sites & > 20 other health systems

Recommendations

- Successful feed forward PROs use design principles
  - Fit PROs into care flow to make it easier for patients and providers to do right thing
  - Co-design with stakeholder input for best end-user utility
  - Educate patients and providers on how to use PROs: providers must pay attention to patient’s data
  - Capture data from other sources to improve utility of information
  - Continuously improve PRO system based on user’s experiences & new technology

Using Patient-Reported Information to Improve Health Outcomes and Health Care Value:

- CASE STUDIES FROM DARTMOUTH, KAROLINSKA AND GROUP HEALTH
- Click Here to Download
Methodological Issues: Method of Administration/Collection & Response

Lewis Kazis, Sc.D
Boston University School of Public Health

OVERVIEW

- Historical Development of Short Form Assessments
- Bridging measurement tools:
  - Legacy/Static measures
  - IRT/CAT measures
- Mixed Mode Approaches
- Missingness / Bias
Timeline
Classical and Modern Test Theory

Classical Test Theory
1960’s
1970’s
1980’s
2000
2010

Modern Test Theory

Binding Framework

Reliability
Validity
Legacy measures
IRT
Item Banks
Facilitate comparison
New Items
Person fit indexes
Precision
CAT
(1) Legacy / Static Measures

- Test score = true score + error score.
- The observed score is item sample dependent, and the statistics are respondent sample dependent.
- Longer tests are more reliable than shorter tests.

(2) Legacy / Static Measures

- Meaningful scale scores are obtained by comparisons of position in a score distribution.
- Interval scale properties are achieved by selecting items that yield normal raw score distributions.
(1) Item Response Theory (IRT) / (CAT)

- A set of generalized linear models that connect observed survey responses to a subject's on an unmeasured underlying latent construct.
- Assume unidimensionality (the scale items solely measure one construct).
- Assume uncorrelated items on a scale.
- Shorter tests can be more reliable than longer tests.

Advantages

- Extensively tested for reliability and validity across multiple settings and populations.
- Fewer resources needed to implement, compared to CAT.
- The expertise to implement them is matured.
- Can be integrated with new technology (internet)

Disadvantages

- The time to complete the instrument is usually longer than CAT.
- Instruments are less flexible to update and recalibrate, compared to CAT.
- Requires larger samples to avoid spurious results.
## (1) Item Response Theory (IRT) / (CAT)

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate person level traits within subset of items.</td>
<td>DIF calculation maybe problematic for multidimensionality assessment (prob. of responding in different cat. vary across different subgroups given equiv. levels of underlying attribute).</td>
</tr>
<tr>
<td>Usually requires smaller sample sizes.</td>
<td>Requires front end technology to implement the instruments.</td>
</tr>
<tr>
<td>Less vulnerable to floor and ceiling effects.</td>
<td>Additional assistance is usually necessary to facilitate successful patient-technology interaction.</td>
</tr>
</tbody>
</table>

## (2) Item Response Theory (IRT) / (CAT)

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Useful in assessing change.</td>
<td>High startup costs.</td>
</tr>
<tr>
<td>Greater precision of measurement.</td>
<td>Software and hardware is commonly proprietary and expensive.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In a number of cases, legacy measures represent the foundation for CAT and item banks development.

Item banks calibration adequately identify problematic legacy wording, enabling the enhancement of legacy measures in terms of reliability and validity.

References.

Facilitate the development of new items to improve existing measures

- Legacy HAQ, SF-36, PF-10, have been improved using PROMIS:
  - Present tense
  - Five-item response categories
  - Improved quality and phrasing

References.
Facilitate comparison across indexes

- Fryback and colleagues found among 5 utility scales, that each measurement identified health in a very similar fashion and are approximately linearly related.

- However, death remains controversial, and its location varied across scales.

Reference.

Modes of Administration

- Face to face interaction.
- Self administration (paper and pencil).
- Telephone.
- Computer-based assessment.
Face to face interaction.

- Overall, responses give a more optimistic picture of health, compared to self-administration.
- White coat effect? It may be related with the positive effect of human involvement (rapport).
- Interviewees overwhelmingly preferred it over the other modes.

Self administration

- Self-report is accurate (fewer sources are variation).
- Lower scores (worse health) are usually reported, compared to face to face modes.
- Less expensive than face to face interviews.
- Anonymity may yield more accurate rates for the “socially undesirable behavior”.
**Telephone**

- Lower response rate compared to personal interviews. However, it costs less than half of the latter.
- Less sensitive to non-response bias, compared to mail surveys.
- Problematic for older adults (higher prevalence of hearing impairments).
- Preferred over self administered surveys for individuals with lower literacy levels.

**Computer-based assessment.**

- Tailored “real time” results, immediately available to users and providers.
- High rates of acceptance, even among interviewees without previous experience with computers.
- The missingness of data may be reduced.
- May capture data more accurately for “socially undesirable behaviors”.
Modes of Administration

- **Electronic vs. Paper instruments.**
  - Both instruments are comparable.
  - A critical review of 56 studies found average correlations exceeding 0.90 between electronic and paper assessments.


Missingness / Bias

- **Lower response rates**
  - Response rates are systematically declining over time.
  - Not enough evidence to determine the potential effects of unit non-response.

- **Response bias**
  - Greatly limits the generalizability of survey findings.
  - Homogeneous populations are less affected by response bias
Missingness / Bias

**Imputation of missing values**
- Missing items
  - Simple mean imputation (should satisfy many conditions first)
  - General imputation methods (e.g. GEE).
- Missing forms
  - multivariate repeated measurements (analysis of variance)
  - modified regression estimates, (MRI estimator).
  - Random effects.

**Discerning Unbiased survey findings**
- Estimating response rates
  - 60% as an acceptability “rule of thumb”.
- Evaluating non response bias
  - More difficult to assess than response bias.
  - However, the representativeness of the sample should be assessed somehow.
- Reporting non response rate.
Summary

- Hybrid approaches necessary that bridge Legacy and IRT/CAT approaches for purposes of application to systems for measurement performance.
- Mixed mode approaches are necessary so that flexibility in the protocols is possible in real world settings.
- Missing data is a fact of life in real world settings and adjustments for missingness is required to adjust for bias in results.
Methodological Issues: Method of Administration/Collection & Response

Lori Frank, PhD
Patient Centered Outcomes Research Institute, Washington, DC

Why Engage?

Guiding Principles for Stakeholder Engagement

- Ask and Prioritize Meaningful Research Questions
- Review Proposals and Conduct Research
- Feedback on Research Impact
- Accelerate Dissemination
Identifying Meaningful Measures

Capturing the Measurement Target

Communicating and Using Results

Relating Performance Measurement to Patient Goals

Guiding Principles for Stakeholder Engagement in Performance Measurement

Why Engage?
Methodological Issues: Selecting Patient-level PROs

Eugene C. Nelson, DSc, MPH
The Dartmouth Institute
Dartmouth-Hitchcock Health

When can general health status measures be utilized & when should condition-specific measures be utilized? Are there any setting specific issues for selection of PROs?

- In general, prudent to use both general and condition-specific measures of functional status and symptoms
  - Examples: Spine, heart failure, total joint replacement of hip and knee, depression
- Imperative to use general health status measures under some conditions
  - When patient has multiple comorbid conditions, e.g., 76.1% of heart failure patients have 2 or more chronic conditions
  - When screening for problems that may be important but can easily go undetected
    - CABG or Spine or AMI or post-partum: screen for mental health problems
    - Annual Wellness Visit or periodic health exam: screen for functional problems and high health risk status
- Settings for PROs use: home, outpatient, inpatient, ECF (subject to patient’s ability to provide data)
What conditions would be most sensitive to measuring changes in patient health status/outcomes? What is the variation in patient-level scores related to clinical interventions (e.g., hip replacement)?

- **Function:** Chronic conditions with large impact on physical, mental and role function such as heart failure, depression, ischemic heart disease, Parkinson’s Disease, low back pain, osteoarthritis, rheumatoid arthritis, etc.
- **Function:** Surgical conditions with large impact on physical, mental and role function such as CABG, TJA, bariatric surgery, spine surgery, etc.
- **Risk:** People at high risk of avoidable death ... health risk status measurement and monitoring using Framingham Index or all cause mortality index such as people with cardio-metabolic syndrome, hypertension, diabetes, hyperlipidemia, and high risk health-related behaviors, etc.

### Table 1. Variation in PROs associated with selected interventions

<table>
<thead>
<tr>
<th>Population</th>
<th>Measure/range</th>
<th>Average Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herniated disk: SPORT</td>
<td>SF-36 PCS / 0-100</td>
<td>44 surg v 32 non-surg</td>
</tr>
<tr>
<td>Stenosis: SPORT</td>
<td>SF-36 PCS / 0-100</td>
<td>17 surg v 17 non-surg</td>
</tr>
<tr>
<td>Spondylolysis: SPORT</td>
<td>SF-36 PCS / 0-100</td>
<td>27 surg v 8 non-surg</td>
</tr>
<tr>
<td>Depression: EBM Protocol</td>
<td>PHQ-9 / 0-27</td>
<td>10 or greater = clinical depression v &lt; 5 = remission or 5-9 = response</td>
</tr>
<tr>
<td>Carpal Tunnel: Trumble Trial</td>
<td>BCTQ/ 1-5</td>
<td>3.1 pre-surg v 1.8 post-surg</td>
</tr>
<tr>
<td>TJA-Hip: UK Knee Society</td>
<td>WOMAC/ 0-100</td>
<td>42 pre-surg v 70 post-surg</td>
</tr>
<tr>
<td>RA: Sweden Registry</td>
<td>DAS / 0-10</td>
<td>5.0 to 2.8 at 1 yr</td>
</tr>
<tr>
<td>Aortic Valve Stenosis: PARTNER Trial</td>
<td>KCCQ/0-100</td>
<td>32 in Transcatheter Aortic Valve Replacement (TAVR) vs 4 in Meds at 1yr</td>
</tr>
</tbody>
</table>
| Angina: COURAGE Trial                    | SAQ Angina Frequency (AF) & QoL/0-100 | AF: 68 vs. 87 baseline v 1 yr post Percutaneous coronary intervention (PCI) QoL: 51 vs. 76 baseline v 1 yr post PCI  
|                                          |                       | AF: 87 in PCI vs. 84 w/meds QoL: 76 in PCI vs. 73 w/meds                      |
Appendix: References

Key Considerations for Incorporating PROs into Electronic Health Records

Uma Kotagal, MSc
Cincinnati Children’s Hospital Medical Center
Overall reduction in symptoms by 40% in most patients as seen on CYBOCS
Recap of Day 1
Overarching Themes
(working draft)

Overarching Theme: Person-Centered

- Terminology: “person” versus “patient”
- Patient experience not just with the healthcare delivery system; includes whether needs are met and linked to other services to improve health
- Patient important outcomes: relevant and meaningful to persons and their families (vs. research)
- PROs must be actionable to persons, providers, policy makers, others. Add to list of essential characteristics
- PROs are an important step towards engaging patients and providers in creating a person-centered environment
Overarching Theme: Accountability

- PROs and state of readiness for purposes of accountability
  - What is the pathway to accountability measures?
  - As measures expand beyond sickness/illness to health/well-being shared accountability will be required beyond the healthcare system

Lessons from the Field

- Person buy-in:
  - Persons feeling “spammed” by survey requests
  - Engage persons in determining what PROs are meaningful to them
- Health Professional buy-in
  - Fitting results into the workflow
  - Knowing what to do with the results
Lessons from the Field

Guiding principles for stakeholder engagement in performance measurement – engage patients at all steps
- Identify meaningful measures
- Capture the measurement target
- Communicating and using results
- Relating performance measures to patient goals

Approaches to Implementation

- Just get started! Let providers innovate.
- Approaches put forth:
  - Initially measure that PROs are collected (e.g., process measure) on relevant patients recognizing outcomes are more meaningful [getting people used to it] NEEDS SHARPENING
  - Usability, feasibility, actionability are paramount to selection of PROs – determine first before implementing into accountability programs
  - Start with focused areas where we have validated measures (e.g., hip heart) and have good evidence on how to improve
  - Generic assessments offering the “biggest bang for the buck”
Themes: Key Methodological Issues

- Missing data
  - Safeguarding against excluding sicker patients
  - Bias introduced by how the tool is administered
  - Engagement strategy needed over time (e.g., response rate)
- Need bridges that combine use of legacy tools and Item Response Theory -- advantages/disadvantages to both suggest hybrid approach

Themes: Key Methodological Issues

- Reconciliation is needed around heterogeneity of multiple approaches (use of different tools & modes of administration) for comparability to make sure they are equivalent
- “Leading” measure versus “lagging” measures (e.g., mortality doesn’t have a guideline)
- Outcomes with high face validity are not required to be based on guidelines (NQF Evidence Task Force)
Themes: Electronic Health Records

- PROs and parsimony
  - Building blocks that can be leveraged for different purposes (e.g., “app” store)
  - Flexible platform (e.g., PROMIS)
  - Infrastructure that exists which can accommodate new tools (e.g. reusable codes)
  - Accommodating multiple styles respectful of ways patients wish to engage with the system

Electronic Health Records

- New reality: Patient will “own” the record and provide and extract information –implications for existing EhRs
- Patient experience is still needed but there are risks putting into EhR (e.g., recrimination)
- IP & copyright issues for codes & instruments (not limited to EhR)
Patient-Reported Outcomes Workshop #1

Breakout Session

July 31, 2012

Framing PROs Within Existing Conceptual Models

Determinants of Health Model

- Genetics & Biometrics
- Physical Environment
- Social Environment
- Lifestyle & Health Behaviors

Patient-Focused Episode of Care Model

PRO Categories Across the Episode:
- HRQOL/Functional Status
- Health-related Behaviors
- Symptom/Symptom Burden
- Patient Experience with health care
AGENDA

- 9:35-9:45  Round Robin Introductions & identify spokesperson
- 9:45-10:30  Discussion of Question #1
  ▫ What characteristics should be used to identify PROs for potential use in performance measures? Will these differ based on the needs of the end-user?
- 10:30-11:15  Discussion of Question #2
  ▫ What existing individual-level PROs have these identified characteristics and are candidates for potential development of performance measures?
- 11:15-11:30  Synthesis and complete templates for report out
Patient-Reported Outcomes Workshop #1

Breakout Session Report Back

July 31, 2012

(working draft)

Report Out: HRQoL/Functional Status

Question #1: Characteristics to Select PROs

- HRQoL/Functional Status
- Adequate measure properties (scientifically and clinically defensible)
  - Conceptual and Measurement Model
  - Reliability
  - Validity (and consideration of validity for proxy response; link to alternative modes)
    - Face validity (clinical/patient sensibility) w/ respect to how relevant it is to patients & clinicians
    - Risk-adjustable
  - Responsiveness
  - Burden
    - Literacy level and cognitive demand
    - Something that be practically incorporated into clinical practices
- Meaningfulness to clinicians and other health professionals, patients, and systems
  - Evidence for usefulness / appropriateness for specific pops
  - Interpretability of Scores
- "Implementability"
  - "Electronicability"/alternate modes
  - Translatable
  - "Game-ability"
  - Unintended consequences
  - "Proprietariness"
**Report Out: HRQoL/Functional Status**

**Question #2: PROs with the Identified Characteristics**

- **HRQoL/Functional Status**
  - Top 3-5 candidates for potential development of performance measures (attached slide color coded **green** for in use)
  
  1. Generic
  2. Disease specific

<table>
<thead>
<tr>
<th>PROs with the Identified Characteristics – Generic</th>
</tr>
</thead>
<tbody>
<tr>
<td>- <strong>AMPAC</strong>&lt;br&gt;  &gt; Mobility and Self-Care</td>
</tr>
<tr>
<td>- <strong>VR12 (med adv)/VR36/VR6D</strong></td>
</tr>
<tr>
<td>- <strong>BRFSS</strong> (three QoL – healthy day items)</td>
</tr>
<tr>
<td>- <strong>PROMIS (adult and children)</strong></td>
</tr>
<tr>
<td>- <strong>EQ5D and HUI</strong> (utility); proprietary</td>
</tr>
<tr>
<td>- <strong>Sickness Impact Profile</strong></td>
</tr>
<tr>
<td>- <strong>Money Follows the Person QoL Scale</strong></td>
</tr>
<tr>
<td>- <strong>Basic and instrumental scale of daily living</strong></td>
</tr>
<tr>
<td>- <strong>Social and productive activity scales</strong></td>
</tr>
<tr>
<td>- <strong>QWB</strong></td>
</tr>
<tr>
<td>- <strong>Child Health questionnaire</strong></td>
</tr>
<tr>
<td>- <strong>CHIP</strong></td>
</tr>
<tr>
<td>- <strong>Kids Screen</strong></td>
</tr>
<tr>
<td>- <strong>FIM (follow-up)</strong></td>
</tr>
<tr>
<td>- <strong>Community Integration questionnaire</strong></td>
</tr>
<tr>
<td>- <strong>SEIQoL</strong></td>
</tr>
<tr>
<td>- <strong>EVGFP</strong></td>
</tr>
<tr>
<td>- <strong>WHO QoL</strong></td>
</tr>
<tr>
<td>- <strong>Penny E</strong></td>
</tr>
<tr>
<td>- <strong>WHO DAS (DALYS)</strong></td>
</tr>
<tr>
<td>- <strong>SF family</strong></td>
</tr>
<tr>
<td>- <strong>Restricted Activity Days</strong></td>
</tr>
</tbody>
</table>
Report Out: Health-Related Behaviors

Question #1: Characteristics to Select PROs

- Health-Related Behaviors
- Synthesis of top 3-5 characteristics to identify PROs for potential use in performance measures

1. *An evidence-based* justification for selection suggesting a measure is actionable to appropriate end-users.
2. Degree of importance to adequately capture the impact of a health-related behavior on a patient
3. Assessment of the level of accountability; individual, culture, environment, resource accessibility, etc
4. Lends itself to a model of shared decision making; engaging patients in their own self-management & goal attainment
Report Out: Health-Related Behaviors
Question #2: PROs with the Identified Characteristics

- Health-Related Behaviors
- Top 3-5 candidates for potential development of performance measures

1. Federally Sponsored Health Surveys
   - BRFSS, NHANES, HOS, ACO CAHPS, Physical Activity FS
2. Commercial Health Risk Appraisals
   - Stay Well, Health Media, U of M,
3. Behavioral Health & Substance Use
   - PHQ-2, CAGE, Audit-C,
4. Specific High Impact Health Related Behaviors
   - How’s Your Health, PROMIS, Smoking Index, Framingham Index

* Inclusion of pediatric category TBD

Report Out: Symptoms and Symptom Burden
Question #1: Characteristics to Select PROs

Symptoms and Symptom Burden
Synthesis of top 3-5 characteristics to identify PROs for potential use in performance measures

- Patient engagement
  - Identify important outcomes to patients
  - Involve in development, testing, use
  - Assure cultural, linguistic, literacy adaptability
- Purpose/Goal
  - Identify end users/stakeholders (patient, caregiver, provider, plan, payor)
  - Specify context of use (disease, population, time horizon, setting, interpretation of results)
  - Articulate conceptual/measurement model
- Actionability
- Measurement properties
  - Content validity
  - Quantitative measurement properties (reliability, construct validity, sensitivity, appropriate recall period)
- Feasibility
  - Consider mode of administration, interoperability with HER
  - Burden to patients, providers, infrastructure requirements
### Report Out: Symptoms and Symptom Burden

**Question #2: PROs to Consider for Measurement**

**Symptoms/Problems and Symptom Burden**
- Pain
  - Worst pain item from the Brief Pain Inventory
- Interference with ADLs/IADLs
- Dyspnea
- Fatigue
- Mood
- Memory
- Sleep disturbance
- Cognitive disturbance
- Mobility
- Nausea/Vomiting
- Constipation
- Diarrhea
- Continence
- Sexual dysfunction
- Appetite loss/anorexia
- Edema
- Body image
- Sensory loss

---

### Report Out: Patient Experience with Healthcare

**Question #1: Characteristics to Select PROs**

- **Goal:** Moving beyond past episode of care: Knowing what happens next, and what patient’s role is - actionable and responsive (i.e., questions that ask about state of patient and ensuing action)

- **Principles:**
  - Continuum of care or longitudinal
  - Not just provider, setting or episode/encounter-specific
  - **Person-centered**
Report Out: Patient Experience with Healthcare

Question #1 Characteristics to Select PROs

- Needs to be actionable that leads to change/improvement by unit of analysis: provider/system as well as the person
- Should be linked to the individual’s goals, which encourages engagement but also flexible enough to account for population
  - Example: PROMIS data bank where items can be selected based on person needs/goals and provider and system needs (setting, etc.)
- Minimize provider/system and patient burden
- Needs to accommodate cultural and language preferences
- Needs to be responsive to individual preferences
  - Include alternative methods of administration (including who, where and when)

Report Out: Patient Experience with Healthcare

Question #2 Existing Tools

- CAHPS (acute/amb care) (NQF-endorsed)
  - provider communication; access to care; timeliness to care; care coordination; patient and family involvement; language access; shared decision-making; care transition; cultural competency; staff helpfulness; experience of environment; alternative medicine; communication about medicine (overlap w/health literacy); pain management
- VA FATE survey (NQF-endorsed)
  - captures episode – family support; care assessment (degree)
Report Out: Patient Experience with Healthcare
Question #2  Existing Tools

- National Core Indicators (long-term care)
  - in use by 35 states; nationwide in 3 years
  - identifies individual responses across multiple areas: family outcomes; individual survey and 3 family surveys; family involvement; health and welfare; therapeutic interventions; medications and incidence; systems issues; safety (home); service coordination; staff stability
  - over 100 individual performance measures (some risk adjusted)
  - 2 sections (consumer section)
  - Who and how they are administering survey also important.
  - How do you balance burden concerns?
    - Background done by case manager; staff input and also receive consumer perspective. Mission of community-based org, consumer engagement throughout

Report Out: Patient Experience with Healthcare
Question #2  Existing Tools

- Patient Activation Tool – shared decision-making; power dynamic; how person felt from encounter; used in chronic care management
  - Concern - measuring output or input?
  - “Self confidence” (suggested domain to measure) – outcome of positive experience. Immediate metric between encounter and outcome
  - Hibbard and Colleagues may have a tool to measure self confidence
Pacific Business Group on Health
- Testing new tool on self care management; health status impact; shared decision-making
- How much consumers pay for services – how they rate their health plan and not provider
- Clinician advocacy on behalf of new patients
- Suggested process measure – did system ask about goals for the visit? Did they respond/fulfill goals?
- Current - Cash and counseling (demo program from CMS) – persons with severe physical disabilities given budget and purchase own services; model worked in 3 states – person make own decisions; current assessment is qualitative, not standardized
- Self directive services – all states offer these options