Guide to Selection of Measures from HealthMeasures

This guide provides information to help you select a person-centered measure from among those provided in PROMIS, Neuro-QoL, ASCQ-Me, and the NIH Toolbox, collectively known as HealthMeasures (see www.healthmeasures.net).

Questions to consider in choosing a HealthMeasure

1. What are the goals and/or aims of the assessment?
2. Should you use a performance-based measure?
3. Do you want to measure global or specific outcomes?
4. Are the patient-centered outcomes primary, secondary or exploratory endpoints?
5. How old is your target population?
6. Do you want to measure disease/condition-specific outcomes or universal (not disease/condition-specific) outcomes?
7. Do you want fixed length outcome measures or dynamic (computer adaptive test, CAT) measures?
8. How reliable, precise and brief does the measure need to be?
9. Is the measure appropriate for my target population?

1. <u>Goals and/or aims of the assessment?</u>
Your measurement goals should drive your measurement selection. It is impractical to measure everything. You will need to prioritize.

If the goal is to improve patient care, you will want to identify the symptoms or outcomes you expect to improve if patient care does, in fact, improve. If the purpose is to facilitate the clinical encounter, you will want to measure the symptoms and outcomes most relevant to patients and whose levels are most likely to drive treatments and referrals.

If you are conducting clinical research, consider what domains would be expected to change with a target intervention. You also should consider what domains should be measured for potential use as covariates.

Whether your goals are clinical, research, or quality improvement, it will be very helpful for you to consider how "proximal" or "distal" a symptom or outcome is to the goal of your research or application. For example, if the goal is to treat anxiety by training patients in meditation, then anxiety is an obvious proximal measure. Physical function, though important to the patient, is distal with respect to the goals of the study.

2. Performance-based measures?
As their name implies, scores on performance-based scales are obtained by observing a patient complete an activity. Tests of hand dexterity and memory/cognitive ability are performance-based measures. These tests are administered by an examiner trained on how to correctly administer the tests. The NIH Toolbox includes a number of performance-based assessments. Examiner training resources include online e-Learning resources, the Training Manual, and in-person workshops (see http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox for more information). Some performance-based measures require specialized equipment.

Though performance-based measures often are referred to as "objective" measures, like all measures, they include measurement error. There are no errorless measures.

If you think a performance-based measure might be appropriate for your research study or clinical application, review the areas, or domains, measured by the NIH Toolbox. The NIH Toolbox includes performance-based measures for children and adults in cognition, motor, and sensation.

3. Global or specific outcome measure?

Global
If you want a measure of overall perceptions of health and health-related quality of life consider the following measures from PROMIS.
- For Adults:
    - 10-item Global Health Scale. Global Physical and Global Mental scores can be obtained using this measure.
    - The PROMIS-29 Profile measure uses four-item scales to measure each of seven domains (Depression, Anxiety, Physical Function, Pain Interference, Fatigue, Sleep Disturbance, and Ability to Participate in Social Roles and Activities).Pain intensity is measured with a single, 0-10 item.
- Children:
    - The PROMIS Global Health 7-item scale produces a single Global Health score. The Global Health 7+2 adds one item each to assess fatigue and pain. Self-report and parent proxy versions are available.
    - The PROMIS-25 Pediatric Profile contains four items each from six PROMIS domains (Depressive Symptoms, Anxiety, Mobility, Pain Interference, Fatigue, and Peer Relationships) and one pain intensity item. A parent proxy version is also available.

Specific
Before selecting a measure of a specific health status domain, consider what you know about your target population or conduct a literature search to learn what areas, or domains, are most relevant. Review the "Intro to PROMIS/Neuro-QoL/ASCQ-Me/NIH Toolbox" pages on the http://www.healthmeasures.net website for a list of available domains.

Some concepts (e.g., working memory) are best assessed using a performance based test administered by an examiner. In these cases, begin with NIH Toolbox measures (see Performance Measures above).

In selecting a PRO measure, consider your target population. Within HealthMeasures are scales that target specific populations including Sickle Cell Disease (ASCQ-Me) and neurological disorders (Neuro-QoL). If your target population has other clinical conditions (or no clinical condition), consider PROMIS.

4. <u>Are the person-centered outcomes primary, secondary or exploratory endpoints?</u>
The primary outcome is the most important outcome that defines the success of the study and is the outcome used to calculate power analyses for sample size determination. Greater precision is needed for primary endpoints. Secondary outcomes are important but may not drive study design decisions. Exploratory outcomes are those not as strongly supported by previous research literature as are the primary and secondary outcomes, but are outcomes worth evaluating for future research.

5. <u>How old is the target population?</u>
An important first question to answer is, "Whose health status do I want to measure?" Is the respondent a child or an adult? If a child, what age? Will respondents be able to report for themselves or will a proxy be needed? (Note: As of 2016, HealthMeasures includes proxy measures for children, but not for adults).

In HealthMeasures, respondents who are 18 years old or older are considered adults. ASCQ-Me only has measures for adults, but the other three systems include measures for different age ranges of children.

6. <u>Disease/condition-specific outcomes or universal outcomes?</u>
Some HealthMeasures instruments are "universal" in their focus. They do not target a particular disease or condition or disability (e.g., PROMIS, NIH Toolbox). Such measures are useful for making comparisons across clinical and non-clinical populations.

Other HealthMeasures are intended for targeted assessment of domains and concerns specific to a disease or condition. ASCQ-Me (sickle cell disease) and Neuro-QoL (neurologic conditions) are examples of such.

<u>Example</u>: A researcher is conducting a study with persons who have Parkinson's disease (PD). She decides that mobility and upper extremity function are important outcomes to assess. Both PROMIS and Neuro-QoL have measures of these domains, but, for the Neuro-QoL instruments, validation data has been obtained in a PD

population. She might also include NIH Toolbox performance tests for gait speed, balance and dexterity.

7. Fixed length or dynamic (computer adaptive test, CAT) outcomes measures?
With fixed length short forms, the same items are administered to every respondent at every time point. These measures can be administered on paper, a computer, or mobile device.

Items also can be administered dynamically using CAT. In this approach, an initial item is administered and, based on the response, a second item is selected and administered. The selected item is the item in the item bank that provides the most information about this person's level of the outcome or symptom being measured. After every response, a computer algorithm updates the score estimate. This process of administering the next best item and updating the estimate continues until a user-defined stopping rule is reached (e.g., stop after 7 items; stop at a given level of precision).

CAT has been shown to reduce test length without loss of precision. However, administering measures by CAT requires access to a computer. To learn more about CATs, watch this educational video series.

8. How reliable, precise and brief does the measure need to be?
Generally, reliability and precision can be increased by asking more questions from an item bank. However, more items exact greater burden on respondents; and, there are diminishing returns from additional items as the scale gets longer. Going from a 1-item scale to a 6–item scale (addition of 5 items) will give you much greater precision. However, going from a 6-item scale to an 11-item scale (also an addition of 5 items) will not provide the same gain in precision.

CAT will almost always be more precise than a fixed short form of the same length. For example, PROMIS CATs can achieve precision that meets standards for individual level assessment in fewer than 6 items on average. A static short form would need more items to obtain the same level of precision.

Typically, the need for precision and the need to limit respondent burden must be balanced. CAT yields brevity and precision but may not be practical in some contexts. Also, there will be some contexts in which you may not need very precise measurement (e.g., large sample studies; studies with large effect expected). Shorter fixed length measures will suffice (e.g., 6 or 8 items) in such cases.

9. Is the measure appropriate for my target population?
In evaluating an instrument, you should look for evidence that it performs well in your target population. If previous qualitative studies have documented the relevance of the

domains and the clarity of the items in samples similar to your target population, no further validity evidence may be necessary. However, if there are hypothesis-driven concerns your target population may experience the outcome differently, it is appropriate to seek more validation evidence. Examples include consulting stakeholders and conducting focus groups or, in-depth interviews.

<u>A FINAL NOTE</u>
Picking an instrument is like picking a life-partner. You can, and you should, weed out the ones that have goals that are different from yours, pay attention to empirical evidence, and remember that you will have to make compromises. The best you can do is find a good match, not a perfect one.