Better Outcomes Measures for Better Rehabilitation Outcomes

Dagmar Amtmann, Ph.D.
dagmara@uw.edu
© University of Washington

Lehman’s day – May 23, 2013
Overview

- New developments in outcomes measurement
- Challenges
- An example of computerized adaptive testing (CAT)
- Examples from clinical practice and research studies
Patient reported outcomes (PROs)

- PROs can be used to monitor patients, fulfill the “review of systems” requirement, aid in making treatment decisions, evaluate treatment effectiveness, facilitate communication and provide documentation to justify services or devices.
- Importance recognized by the NIH, the FDA and by ObamaCare that funds Patient Centered Outcomes Research Institute (PCORI).
Many measures of the same domain
Varying definitions and quality of scales
Difficult to compare and combine data

... across studies
... across conditions

Scales are often long and all items have to administered to all persons
The Tower of Babel (Brueghel, 1563)
Change of paradigm in measurement

From Classical Test Theory
One size fits all

To
Item Response Theory
Custom Administration
Brief, Precise and Flexible
Measurement initiatives to help find the common language

- **Self-report**
  - NIH funded
    - PROMIS
      nihpromis.org
    - Neuro-QoL
      neuroQOL.org

- **NIDRR funded**
  - SCI-QOL
  - TBI-QOL

- **Performance-based**
  - (NINDS funded)
    - NIH Neurotoolbox
      (cognitive, motor, sensory and emotional health)
      nihtoolbox.org
A little bit of vocabulary…

A **domain** is the specific feeling, function, or perception you want to measure.

Our domain of interest for the demo: 
Physical function
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 Metric

• T Score
  — Mean = 50
  — SD = 10

• Referenced to the US General Population
Item response theory (IRT) uses the same logic as an experienced interviewer

- Takes what is known about
  - the items (e.g., item difficulty), and
  - the pattern of responses to the items
- Makes an estimate of a person’s most likely trait level.
- This is accomplished with a mathematical “model”—an IRT model
IRT in plain English

- How likely an individual is to respond in a particular way to a particular item depends on:
  - How much they have of the trait being measured and,
  - What the characteristics of the item are (e.g., how difficult).
Physical Functioning Item Bank

Are you able to get in and out of bed?
Are you able to stand without losing your balance for 1 minute?
Are you able to walk from one room to another?
Are you able to walk a block on flat ground?
Are you able to run or jog for two miles?
Are you able to run five miles?
Computerized Adaptive Testing (CAT)?

- CAT integrates the advances in measurement theory and the power of computer technology to administer a PRO instrument that selects questions on the basis of a patient’s response to previously administered questions.

- Most informative questions are carefully selected so that we can estimate a person’s standing on a domain (e.g., physical functioning, depression) with the minimal number of questions without a loss in measurement precision.
Demonstration of Computerized Adaptive Testing (CAT)
Are you able to keep up with others when walking?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
Are you able to walk while carrying a shopping basket in one hand?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
Are you able to take 2 steps backwards?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
Are you able to walk between rows of occupied seats like those in a theater or church?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
CAT Score

Example: low mobility

- Theta (logit) = -1.13
- T-score = 38.7
Are you able to keep up with others when walking?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
Are you able to walk up and down steep stairs in a crowded stadium?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
Are you able to walk in soft sand?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
Are you able to hike about 2 miles on uneven surfaces, including hills?

- Without any difficulty
- With a little difficulty
- With some difficulty
- With much difficulty
- Unable to do
CAT Score

Example: high mobility

- Theta (logit) = 3.16
- T-score = 81.6
Advantages of Item Banks

• **Computerized Adaptive Testing (CAT)**
  - Dynamic testing
  - Median 4 items, but precise score

• **Fixed Length Instruments (SFs)**
  - Targeted to sub-populations (e.g., uni- or bi-lateral amputation, below or above the knee amputation)
  - Targeted to high or low mobility
  - Different lengths (for instance 2 items for a quick screen, 10 items as a primary outcome in a clinical trial)
Item Banks Allow Multiple Delivery Platforms

Modes of Assessments

- Smart phones
- Portable devices
- Internet Administered
- Self Administered
- Interactive Voice Recognition
- Targeted Short Forms
Fatigue Item Bank

- Hypnosis Study (Chronic pain)
  - Items 1-10
  - Exercise & depression trial (PPS, MS, MD, SCI)  
    - Items 2, 4, 9, 13

- ESI study (Back pain)
  - CAT

- Symptom self-management trial (MS)
  - Items 6-12
  - Prosthetic knee study (amputation)  
    - Items 1-5

Same metric, same meaning
Examples of use in clinical practice

- Screening (e.g., depression)
- Monitoring of patients (any new/worse symptoms, how is the new medication working)
- Review of symptoms
- Treatment effectiveness
- Population monitoring
MS symptom profiles (N=594)
MS sample by age

- Pain impact
- Wake function
- Sleep function
- Depression
- Anxiety
- Fatigue
- Global mental
- Global physical


-0.87 -1.97 -0.52 -1.73 -0.87 -0.29 -0.24 -1.74 -3.41 -2.32 -2.20 -1.74

5.57 3.23 1.97 1.73 3.41 2.32 1.73 3.41 5.31 3.41 3.41 1.73

7.03 3.34 2.32 2.32 4.06 4.06 2.32 4.06 5.14 4.06 4.06 2.32
MS symptoms Female compared to Male

- Pain impact
- Wake function
- Sleep function
- Anxiety
- Depression
- Fatigue
- NeuroQoL General concerns
- NeuroQoL Executive functioning
- Global mental

- P7 female norm minus PROMIS female norm
- P7 male norm minus PROMIS male norm

Values:
- Pain impact: Female 4.09, Male 5.71
- Wake function: Female -0.22, Male 1.74
- Sleep function: Female 1.42, Male 4.29
- Anxiety: Female 0.60, Male 0.66
- Depression: Female 0.78, Male 1.04
- Fatigue: Female 7.15, Male 7.83
- NeuroQoL General concerns: Female -7.81, Male -6.93
- NeuroQoL Executive functioning: Female -6.72, Male -6.60
- Global mental: Female -5.27, Male -4.08
Pre- and post-ESI scores

- Pain Interference
  - Baseline: 63.8
  - 90 days: 58.5
- Pain Behaviors
  - Baseline: 59.7
  - 90 days: 56.5
- Fatigue
  - Baseline: 56.6
  - 90 days: 53.2
- Depression
  - Baseline: 53.0
  - 90 days: 50.6
- Anxiety
  - Baseline: 55.5
  - 90 days: 52.8
- Sleep Disturbance
  - Baseline: 56.0
  - 90 days: 52.4
- Physical Function
  - Baseline: 40.9
  - 90 days: 37.8
- Social Roles
  - Baseline: 46.4
  - 90 days: 42.4
- Discretionary Social Activities
  - Baseline: 45.8
  - 90 days: 41.5

All changes statistically significant (p < 0.001)
The graph illustrates the health profiles of persons with lower limb loss. The x-axis represents the T-Score Difference (Sample - US Norms), with values ranging from -10.0 to 10.0. The y-axis lists various health domains: Anxiety, Sleep disturbance, Depression, Pain interference, Fatigue, Social role satisfaction, and Physical function.

- **Anxiety**: Better than general population (T-score: -0.59)
- **Sleep disturbance**: Better than general population (T-score: -0.61)
- **Depression**: Better than general population (T-score: -0.71)
- **Pain interference**: Worse than general population (T-score: 4.83)
- **Fatigue**: Better than general population (T-score: -1.53)
- **Social role satisfaction**: Worse than general population (T-score: -2.05)
- **Physical function**: Worse than general population (T-score: -8.15)

The legend indicates that green bars represent scores better than the general population, while red bars represent scores worse than the general population.
Clinics routinely collecting PROs

- UW Pain Clinic
- Cleveland Clinic
- Rheumatology Clinic @ Cincinnati Children’s Hospital & MC
- Robert H. Lurie Comprehensive Cancer Center at NWU
Routine collection of patient-reported outcomes in an HIV clinic setting: the first 100 patients.
Crane HM, Lober W, Webster E, Harrington RD, Crane PK, Davis TE, Kitahata MM.
The Department of Medicine, University of Washington, Seattle, Washington, USA. hcrane@u.washington.edu

Abstract

BACKGROUND: Information from patient-reported outcomes (PROs) can enhance patient-provider communication and facilitate clinical research. However, there are barriers to collecting PROs within a clinic. Recent technological advances may help overcome these barriers. We examined the feasibility of using a web-based application on tablet PCs with touch screens to collect PROs in a busy, multi-provider, outpatient HIV clinical care setting.

METHODS: Patients presenting for routine care were asked to complete a touch-screen-based assessment containing 62 to 111 items depending on patient responses. The assessment included instruments measuring body morphology abnormalities, depression, symptom burden, medication adherence, drug/alcohol/tobacco use, and health-related quality of life.

RESULTS: Of 136 patients approached to participate in the study, 106 patients (78%) completed the assessment, 6 (4%) started but did not complete it, and 24 (18%) refused. Of those who completed the assessment, the mean age was 48 years, and 29% reported a history of injection drug use. The median time to complete the assessment was 12 minutes. The prevalence of lipoatrophy was 51%, the prevalence of lipohypertrophy was 69%, and the prevalence of moderate or severe depression was 51%. We found that 25% of those receiving highly active antiretroviral therapy noted missing a dose of their antiretroviral medications in the prior 4 days.

CONCLUSIONS: Collection of PROs using touch-screen-based, internet technology was feasible in a busy HIV clinic. We found a high prevalence of body morphology abnormalities, depression, and poor adherence. Touch-screen-based collection of PROs is a promising tool to facilitate research and clinical care.

PMID: 17266562 [PubMed - indexed for MEDLINE]
Screening for depression in HIV

**Patient reported outcomes in routine care: advancing data capture for HIV cohort research.**

Department of Internal Medicine, Division of Infectious Diseases, University of Alabama at Birmingham, AL 35294-2050, USA. mkozak@uab.edu

**Abstract**

**INTRODUCTION:** Computerized collection of standardized measures of patient reported outcomes (PROs) provides a novel paradigm for data capture at the point of clinical care. Comparisons between data from PROs and Electronic Health Records (EHR) are lacking. We compare EHR and PRO for capture of depression and substance abuse and their relationship to adherence to antiretroviral therapy (ART).

**METHODS:** This retrospective study includes HIV-positive patients at an HIV clinic who completed an initial PRO assessment April 2008-July 2009. The questionnaire includes measures of depression (PHQ-9) and substance abuse (ASSIST). Self-reported ART adherence was modeled using separate logistic regression analyses (EHR vs PRO).

**RESULTS:** The study included 782 participants. EHR vs PRO diagnosis of current substance abuse was 13% (n = 99) vs 6% (n = 45) (P < .0001), and current depression was 41% (n = 317) vs 12% (n = 97) (P < .0001). In the EHR model, neither substance abuse (OR = 1.25; 95% CI = 0.70-2.21) nor depression (OR = 0.93; 95% CI = 0.62-1.40) was significantly associated with poor ART adherence. Conversely, in the PRO model, current substance abuse (OR = 2.78; 95% CI = 1.33-5.81) and current depression (OR = 1.93; 95% CI = 1.12-3.33) were associated with poor ART adherence.

**DISCUSSIONS:** The explanatory characteristics of the PRO model correlated best with factors known to be associated with poor ART adherence (substance abuse; depression). The computerized capture of PROs as a part of routine clinical care may prove to be a complementary and potentially transformative health informatics technology for research and patient care.
Feasibility and construct validity of PROMIS and "legacy" instruments in an academic scleroderma clinic.


Division of Rheumatology, University of Michigan, Ann Arbor, MI 48109, USA. khannad@med.umich.edu

Abstract

OBJECTIVE: The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS) roadmap initiative is a cooperative group program of research designed to develop, evaluate, and standardize item banks to measure patient-reported outcomes relevant across medical conditions. The objective of the current study was to assess feasibility and evaluation of the construct validity of PROMIS item banks versus legacy measures in an observational study in systemic sclerosis (SSc). We hypothesized that the PROMIS item banks can be administered in a clinical setting if there is adequate staff support without disrupting the flow of clinic.

METHODS: Patients with SSc in a single academic center completed computerized adaptive test (CAT) administered PROMIS item banks during the clinic visit and legacy measures (using paper and pencil). The construct validity of PROMIS items was evaluated by examining correlations with corresponding legacy measures using multitrait-multimethod analysis.

RESULTS: Participants consisted of 143 SSc patients with an average age of 51.5 years: 71% were female and 68% were white. The average number of items completed for each CAT-administered item bank ranged from 5 to 8 (69 CAT items per patient), and the average time to complete each CAT-administered item bank ranged from 48 seconds to 1.9 minutes per patient (average time = 11.9 minutes/per patient for 11 banks). All correlations between PROMIS domains and respective legacy measures were large and in the hypothesized direction (ranged from 0.61 to 0.82).

CONCLUSION: Our study supports the construct validity of the CAT-administered PROMIS item banks and shows that they can be administered successfully in a clinic with support staff. Future studies should assess the feasibility of PROMIS item banks in a busy clinical practice.

Copyright © 2012 International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. All rights reserved.
RCT of effects of PROs

Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial.

Harvard Medical School, Boston, MA, USA. donna_berry@dfci.harvard.edu

Abstract

PURPOSE: Although patient-reported cancer symptoms and quality-of-life issues (SQLIs) have been promoted as essential to a comprehensive assessment, efficient and efficacious methods have not been widely tested in clinical settings. The purpose of this trial was to determine the effect of the Electronic Self-Report Assessment-Cancer (ESRA-C) on the likelihood of SQLIs discussed between clinicians and patients with cancer in ambulatory clinic visits. Secondary objectives included comparison of visit duration between groups and usefulness of the ESRA-C as reported by clinicians.

PATIENTS AND METHODS: This randomized controlled trial was conducted in 660 patients with various cancer diagnoses and stages at two institutions of a comprehensive cancer center. Patient-reported SQLIs were automatically displayed on a graphical summary and provided to the clinical team before an on-treatment visit (n = 327); in the control group, no summary was provided (n = 333). SQLIs were scored for level of severity or distress. One on-treatment clinic visit was audio recorded for each participant and then scored for discussion of each SQLI. We hypothesized that problematic SQLIs would be discussed more often when the intervention was delivered to the clinicians.

RESULTS: The likelihood of SQLIs being discussed differed by randomized group and depended on whether an SQLI was first reported as problematic (P = .032). Clinic visits were similar with regard to duration between groups, and clinicians reported the summary as useful.

CONCLUSION: The ESRA-C is the first electronic self-report application to increase discussion of SQLIs in a US randomized clinical trial.

Comment in

Supporting clinical practice decisions with real-time patient-reported outcomes. [J Clin Oncol. 2011]
Perspectives of patients on the utility of electronic patient-reported outcomes on cancer care. [J Clin Oncol. 2011]
The Electronic Self Report Assessment for Cancer and Self-Care Support: Results of a Multi-Center Randomized Trial (Donna L Berry at al submitted)

- 752 eligible ambulatory adult participants with various diagnoses were randomized to symptom/quality of life (SxQOL) screening at 4 time-points (control) versus screening, targeted education, communication coaching and the opportunity to track/graph SxQOL over time (intervention). A summary of the participant-reported data was delivered to clinicians at each time-point in both groups.

- Results: significant difference between study groups in mean SDS-15 score change from baseline to end of study: 1.27 (SD=6.7) in the control group (higher distress) versus -0.04 (SD=5.8) in the intervention group (lower distress) (t=2.40, p=0.017). The benefit of intervention was strongest in those over 50 years (p=0.04).
**PROMIS Scores**

Scores are reported in PERCENTILES and compared to a sample matched to the US 2000 Census on age, race/ethnicity, and sex. Higher scores indicate BETTER HEALTH.

<table>
<thead>
<tr>
<th>Date</th>
<th>Score</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 May 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>05 Aug 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 Nov 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 Feb 12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Treatment History**

### Healthcare providers seen in the past 6 months:

- General Practice: 1
- Medical Specialists: 4
- Psychologists, Psychiatrists, other mental health professionals: 2
- Allied health professionals: 1
- Complementary and alternative healthcare professionals: 0

### Treatment modalities and effectiveness, in the past 6 months:

<table>
<thead>
<tr>
<th>Treatment Modalities</th>
<th>Effective?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise, physical therapy or</td>
<td>Yes</td>
</tr>
<tr>
<td>occupational therapy</td>
<td></td>
</tr>
<tr>
<td>Physical modalities such as heat,</td>
<td>Moderately</td>
</tr>
<tr>
<td>massage, or TENS</td>
<td></td>
</tr>
<tr>
<td>Behavioral treatment (CBT, relaxation,</td>
<td>Yes</td>
</tr>
<tr>
<td>distraction, etc.)</td>
<td></td>
</tr>
<tr>
<td>Non-opioid, non-steroidal anti-</td>
<td>Yes</td>
</tr>
<tr>
<td>inflammatory medications</td>
<td></td>
</tr>
<tr>
<td>Non-opioid, non-steroidal, neuropathic</td>
<td>Very</td>
</tr>
<tr>
<td>pain medications</td>
<td></td>
</tr>
<tr>
<td>Alternative therapies such as</td>
<td>Yes</td>
</tr>
<tr>
<td>acupuncture, hypnosis, yoga or</td>
<td></td>
</tr>
<tr>
<td>meditation</td>
<td></td>
</tr>
</tbody>
</table>

**Opioid Utilization Screener**

- Currently taking opiates/opioid/narcotics? Yes
- How long: Pain relief
  - Pain relief: 26 months
  - “Bad days” in past month: Good
  - 3-5
New Challenges

- Integration with medical records
- Point of care access
- Logistics of administration in clinics
  - Workflow disruption
  - Technology (maintenance, security, disinfection between uses)
  - Support for patients/staff time
- Presenting data in easily interpreted and actionable formats
THANK YOU

dagmara@uw.edu