Health-Related Quality of Life - 2011

John E. Ware, Jr., PhD, Professor and Chief
Measurement Sciences Division, Department of Quantitative Health Sciences
University of Massachusetts Medical School, Worcester, MA

Introduction to the Principles and Practice of Clinical Research
National Institutes of Health - Warren G. Magnuson Clinical Center Building 10
Lipsett Amphitheater, Bethesda, MD, January 31, 2011
What Are The Advances?

• Standardization (concepts & metrics)
• Better measures
• Adaptive (dynamic) survey administration
• Internet and other technologies
Quality of Life

- Community
- Education
- Family Life
- Friendships
  - Health
- Housing
- Marriage
- Nation
- Neighborhood
- Self
- Standard of Living
- Work

Source: Campbell, 1981
World Health Organization
Definition of Health

“Health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity”

WHO, 1948
Health is Measured in Terms of:

- Bodily structure & function
- Specific symptoms
- What you do/are able to do – functioning
- How you feel – subjective ill- and well-being (+ and -)
- What you say it is – personal evaluation

Sources: Understanding Health Outcomes Educational Series
Continuum of Disease-specific and Generic Health Measures - Arthritis

X-ray, Disease Progression

Clinical Markers

(1)

Arthritis Symptoms

Have you had the following symptoms:
- Joint pain, swelling, burning sensation:
  - Almost every day
  - Several days a week
  - A few days a month
  - Not at all

(2)

Arthritis Impact

How much does your arthritis limit your usual activities or enjoyment of everyday life?
- Not at all
- A little
- Moderately
- Extremely

(3)

Generic Health

In general, would you say your health is:
- Excellent
- Very good
- Good
- Fair
- Poor

(4)

Impact of Disease-specific Problems

Health-related QOL (HR-QOL)

Adapted from: Wilson and Cleary, JAMA, 1995
Ware, Annual Rev. Pub. Health, 1995
## Summary of Content/ Concepts for Widely-Used Generic Health Surveys

<table>
<thead>
<tr>
<th>Concepts and Characteristics</th>
<th>Psychometric</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SIP HIE NHP QLI COOP DUKE MOS FWBP SF-36 MOS PROMIS QWB EURO QOL HUI SF-6D</td>
<td></td>
</tr>
<tr>
<td><strong>CONCEP</strong>TS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning (-)</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Social functioning (-)</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Role functioning (-)</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Psychological distress (-)</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Health perception (gen’l)</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Pain (-)</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Vitality (-)</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Psychological well-being</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Sleep</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Quality of life</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Reported health transition</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

**SIP** = Sickness Impact Profile (1976)

**HIE** = Health Insurance Experiment surveys (1979)

**NHP** = Nottingham Health Profile (1980)

**QLI** = Quality of Life Index (1981)

**COOP** = Dartmouth Function Charts (1987)

**DUKE** = Duke Health Profile (1990)

**MOS FWBP** = MOS Functioning & Well-Being Profile (1992)

**MOS SF-36** = MOS 36-Item Short-Form Health Survey (1992)


**QWB** = Quality of Well-Being Scale (1973)

**EUROQLQOL** = European Quality of Life Index (1990)

**HUI** = Health Utility Index (1996)

**SF-6D** = SF-36 Utility Index (2002)

Source: Adapted from Ware, 1995
Visit the PROMIS Website

www.nihpromis.org
Anatomy of a Survey Item

Recall period

During the past 4 weeks, how often did your ATTRIBUTION limit your ability to do your everyday activities such as work, school or chores?

Item stem

• Never
• Rarely
• Sometimes
• Often
• Very often

Domain/concept

Attribution

Response categories
SF-36 Health Survey Measurement Model

<table>
<thead>
<tr>
<th>Items</th>
<th>Scales</th>
<th>Summary Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a. Vigorous Activities</td>
<td>Physical Functioning (PF)</td>
<td>Physical</td>
</tr>
<tr>
<td>3b. Moderate Activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c. Lift, Carry Groceries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3d. Climb Several Flights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3e. Climb One Flight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3f. Bend, Kneel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3g. Walk Mile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3h. Walk Several Blocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3i. Walk One Block</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3j. Bathe, Dress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a. Cut Down Time</td>
<td>Role-Physical (RP)</td>
<td></td>
</tr>
<tr>
<td>4b. Accomplished Less</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4c. Limited in Kind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4d. Had Difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Pain-Magnitude</td>
<td>Bodily Pain (BP)</td>
<td>Mental</td>
</tr>
<tr>
<td>8. Pain-Interfer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. EVGFP Rating</td>
<td>General Health (GH)*</td>
<td></td>
</tr>
<tr>
<td>11a. Sick Easier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11b. As Healthy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11c. Health To Get Worse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11d. Health Excellent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9a. Pep/Life</td>
<td>Vitality (VT)*</td>
<td></td>
</tr>
<tr>
<td>9b. Energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9g. Worn Out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9h. Tired</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Social-Extent</td>
<td>Social Functioning (SF)*</td>
<td></td>
</tr>
<tr>
<td>10. Social-Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a. Cut Down Time</td>
<td>Role-Emotional (RE)</td>
<td></td>
</tr>
<tr>
<td>5b. Accomplished Less</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5c. Not Careful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9b. Nervous</td>
<td>Mental Health (MH)</td>
<td></td>
</tr>
<tr>
<td>9c. Down in Dumps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9d. Peaceful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9f. Blue/Sad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9h. Happy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant correlation with other summary measure.
There is More to the Continuum

1. Clinical Markers
2. Specific Symptoms
3. Impact of Disease-specific Problems
4. Generic Functioning, Well-being, and Evaluation
Predictive Validity: HR-QOL is one of the Best Predictors

Impact of Disease-specific Problems

Generic Functioning, Well-being and Evaluation

Health-Related QOL (HR-QOL)

Future health
Inpatient expenditures
Outpatient expenditures
Job loss
Response to treatment
Return to work
Work productivity
Mortality

(3) (4)
Methods:

- Standardization (norm-based scoring)
- Item response theory (IRT)
- Computerized adaptive testing (CAT)
- Internet and handhelds
Improving the Physical Function “Ruler”

1980 Old “Ruler”
> 75% @ Ceiling

1990 NEW “Ruler”
> 30% @ Ceiling

2008 BETTER “Ruler”
< 3% @ Ceiling

Vigorous Activities, Not limited

Norm

Climbing several flights of stairs

Walk one hundred yards

Bathing or dressing, Limited a little

Mean = 50
SD = 10
International Quality of Life Assessment (IQOLA) Project

• **Standardized** questionnaires and scoring methods

• **Culturally sensitive** questionnaires
Item Response Model Calibrations for Physical Function Items, Seven Countries

Denmark (82)  
Italy (82)  
UK (81)  
Netherlands (79)

Sweden (82)  
US (80)  
Germany (79)

Vigorous activities

Denmark (32)  
Italy (25)  
UK (24)  
Netherlands (22)

Sweden (31)  
Germany (29)  
US (24)

Bathing

Source: Raczek, Ware, Bjorner et al, JCE, 1998
Short-Form Surveys and Ceiling Effects

Measuring Too Low - Ceiling Effect
IRT-Based Item Bank Combining HAQ and MOS Items Yielded a Better Scale

Some Thermometers Focus on a Very Narrow Range

Cooking Thermometer

130–190 °F
54–88 °C
Example: Cross-Calibrating Celsius and Fahrenheit

Temperature

F° = 98.6
C° = 37.0

- Normal Human Blood
- Shirt Sleeve Weather
- Water Freezes
Cross-Calibration Makes Scores Comparable and Interpretable

<table>
<thead>
<tr>
<th>Scales</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDI</td>
<td>16</td>
<td>43</td>
<td>73</td>
<td>91</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>HIMQ</td>
<td>74</td>
<td>53</td>
<td>31</td>
<td>17</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>MIDAS</td>
<td>58</td>
<td>28</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MSQ</td>
<td>31</td>
<td>53</td>
<td>79</td>
<td>92</td>
<td>96</td>
<td>99</td>
</tr>
<tr>
<td>DYNHA-5 (+)</td>
<td>23</td>
<td>32</td>
<td>41</td>
<td>51</td>
<td>58</td>
<td>66</td>
</tr>
</tbody>
</table>

Theta (θ) [Best Possible Estimate]

Note: Direction of scoring shown with arrows  
Source: Ware, Bjorne & Kosinski, *Medical Care*, 2000
Standardization: Scoring

Scoring Software

Static Short Forms

Dynamic Short Forms

“Improvements in Short Form Measures of Health Status,”
_J. Clinical Epidemiology_, 2008
Solution: Assess Health Dynamically

Patient scores here

CAT = Computerized Adaptive Testing
1. Begin with initial score estimate

2. Select & present optimal survey item

3. Score response

4. Re-estimate health score and confidence interval

5. Is stopping rule satisfied

6. End scale assessment

7. End of battery?

8. Administer next scale

9. Stop

Source: Adapted from Wainer et al. (2000)
Dynamic Assessments Match Questions to Each Patient’s Level

- **Severe**
  - 80
  - 70
  - 60

- **Moderate**
  - 50
  - 40
  - 30

- **Mild**
  - 20
  - 10
First Question
Noisy Score Estimate (+/- 15)

First Item
Score = 62
+/- 15

Severe

Moderate

Mild
Second Question: Standard Error Reduced by One Third

Score = 64

+/- 10
Third Question: Standard Error Cut in Half

Third Item Score = 63 +/- 7
Fourth Question: Standard Error Cut by Two Thirds

Severe

Moderate

Mild

Fourth Item Score = 62 +/- 5
A Promising Solution in 1999: CAT-Based Health Assessment

“Ceiling Effect”

Criterion Score

r = 0.536
N = 1016

3 SD units

No Disability

Criterion Score

r = 0.938
N = 1016

Static 5-Item Headache Pain Measure

Dynamic 5-Item Headache Pain Measure

Practical Implications of CAT in Health Assessment

MEDICAL CARE
Volume 38, Number 9, Supplement II, pp II-73–II-82
©2000 Lippincott Williams & Wilkins, Inc.

Practical Implications of Item Response Theory and Computerized Adaptive Testing

A Brief Summary of Ongoing Studies of Widely Used Headache Impact Scales

JOHN E. WARE, JR, PHD,*† JAKOB B. BJORNER, MD, PHD,*† AND MARK KOSINSKI, MA*

We have the potential to substantially advance the field of health status assessment by constructing and calibrating questionnaires based on item response theory (IRT) and administering them using computerized adaptive methods. This opportunity could untreated. It was hoped that an accurate and user-friendly report of headache impact would be useful to patients and those who treat them. To benefit as much as possible from prior work and to maintain comparability of scores with current
Performance of 5-item CAT Scores Confirmed in NIH-Sponsored Studies

- Mental Health: $r = 0.98$, $N = 2,753$
- Headache Disability: $r = 0.94$, $N = 1,016$
- Pediatric Disability: $r = 0.96$, $N = 263$
- Chronic Kidney Disease: $r = 0.95$, $N = 1,846$
- Diabetes Impact: $r = 0.93$, $N = 100$
- Post Acute Rehabilitation: $r = 0.95$, $N = 485$
What are the Advantages of Dynamic Assessments?

- More accurate risk screening
- Reliable enough to monitor individual outcomes
- Brevity of a short form – 90% reduction in respondent burden
- Elimination of “ceiling” & “floor” effects
- Can be administered using various data collection technologies
- Markedly reduced data collection costs
- Monitor data quality in real time
Matching Methods to Applications

Population Surveys
- Noisy Individual Classification
- Single-Item

Group-Level Studies
- Multi-Item Scale

Patient-Level Assessments
- "Item Bank" (CAT Dynamic)
- Very Accurate Individual Classification
- Most Functionally Impaired
Dynamic Short-Forms are More Precise

“Improvements in Short Form Measures of Health Status,”

*J. Clinical Epidemiology*, 2008
HR-QOL Validation Strategies

- Diagnosis
- Disease severity
- Responders
- Treatments

Gold Standard

HR-QOL

- Work productivity
- Costs of care
- Mortality
- Self-evaluated health

Other Measures & Methods

Using Comprehensive Item Banks as Criteria in Validating Physical Function Measures

**Rheumatoid Arthritis**

- **PF-2 (“Static”)**: Reliable 0.90, CAT Range: 0.75
- **PF-10 (“Static”)**: Reliable 0.90, CAT Range: 0.75
- **PF CAT-10**: Reliable 0.95, CAT Range: 0.95

Gold Standard (136-Item Bank, N=17,726)

HR-QOL Validation Strategies

Gold Standard

Clinical Causes
- Diagnosis
- Disease severity
- Clinical endpoint
- Treatment

HR-QOL

Other Measures & Methods

Economic & Social Consequences
- Work productivity
- Costs of care
- Mortality
- Self-evaluated health

Standardizing Metrics Across Applications

- Chronic Disease Registry
- Population Health Surveys

Chronic Disease Registry:
- Congestive Heart Failure
- Chronic Lung Disease
- Chronic Kidney Disease
- Diabetes Type II
- Asthma

MOS Physical Component Summary (PCS)
(Mean = 50, SD = 10)
Are Generic HR-QOL Measures Responsive?

SF-36 agreed with primary endpoint (across Rx):

- 219 of 253 RCTs
- 86.6%

X-ray, Disease Progression

Clinical Markers

Specific Symptoms

Impact of Disease-specific Problems

Generic Health

In general, would you say your health is:

- Excellent
- Very good
- Good
- Fair
- Poor

Reference: Ware JE and Frendl D  “Systematic review of the responsiveness of SF-36 Health Survey measures to efficacious pharmaceutical therapies in published double-blind randomized controlled trials.” Presentation at the 13th Annual European ISPOR Congress, Prague Czech Republic, November 6-9, 2010
Internet: Sampling, Data Collection and Reporting

www.amIhealthy.com

www.asthmacontroltest.com


Health Outcomes Research Using Handhelds
Final Comments

• Continuum of outcomes (specific-generic)
• Standardization of concepts and metrics
• Advances in psychometric methods
• Advances in data collection methods
• Comprehensive validation
Thank You!

For more information:  john.ware@umassmed.edu