Health-Related Quality of Life - 2010

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, February 1, 2010
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

• What you are able to do – functioning

• How you feel – distress & well-being

• What you say it is – personal evaluation

Sources: Understanding Health Outcomes Educational Series
What Do We Need to Measure Health Outcomes in the 21st Century?

- Outcomes that matter to patients
- Practical measures
- Coverage of a wide range
- Greater precision
- Comparability of scores
- Ease of interpretation
Continuum of Disease-specific and Generic Health Measures - CKD

Serum Creatinine

Clinical Markers

CKD Symptoms

Specific Symptoms

Over the last 4 weeks I have had the following symptoms: itching, muscle cramps, fatigue:
Almost every day
Several days a week
A few days a month
Not at all

How much did your lung/respiratory problems limit your usual activities or enjoyment of everyday life?
Not at all
A little
Moderately
Extremely

Impact of Disease-specific Problems

Another name: Patient-Reported Outcomes (PROs)

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

Generic Functioning, Well-being and Evaluation

Health-related QOL

Adapted from: Wilson and Cleary, JAMA, 1995
Ware, Annual Rev. Pub. Health, 1995
# Summary of Information About Widely-Used General Health Surveys

**Source:** Ware JE, Jr.: Standards for validating health measures: Definition and content. *Journal of Chronic Diseases* 1987;40:473-80.  
Adapted from Ware, 1995

<table>
<thead>
<tr>
<th>Concepts and Characteristics</th>
<th>SIP</th>
<th>HIE</th>
<th>NHP</th>
<th>QLI</th>
<th>COOP</th>
<th>DUKE</th>
<th>MOS FWBP</th>
<th>MOS SF-36</th>
<th>QWB</th>
<th>EUROQOL</th>
<th>HUI</th>
<th>SF-6D</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCEPTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological distress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health perceptions (general)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain (bodily)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy/fatigue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological well-being</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported health transition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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 and Well-Being Profile (1992)  
- **MOS SF-36** = MOS 36-Item Short-Form Health Survey (1992)  
- **QWB** = Quality of Well-Being Scale (1973)  
- **EUROQOL** = European Quality of Life Index (1990)  
- **HUI** = Health Utility Index (1996)  
- **SF-6D** = SF-36 Utility Index (Brazier, 2002)

**NIH Roadmap PROMIS Initiative:**  
- Fatigue  
- Negative affect  
- Pain  
- Physical Function  
- Social/role activity  

[www.nihpromis.org](http://www.nihpromis.org)
SF-36 Health Survey Measures

Physical
- Physical Functioning (PF)
- Role-Physical (RP)
- Bodily Pain (BP)

Mental
- General Health (GH)*
- Vitality (VT)*
- Social Functioning (SF)*
- Role-Emotional (RE)
- Mental Health (MH)

Utility Index (Brazier et al., 2002)

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

Clinical Markers (1) → Specific Symptoms (2) → Impact of Disease-specific Problems (3) → Generic Functioning, Well-being and Evaluation (4)
Prediction and Risk Management: HRQOL is one of the Best Predictors

Health-Related QOL (HR-QOL)

Impact of Disease-specific Problems

Generic Functioning, Well-being and Evaluation

Future health
Inpatient expenditures
Outpatient expenditures
Job loss
Response to treatment
Return to work
Work productivity
Mortality
What Do We Need to Measure Health Outcomes in the 21st Century?

- Outcomes that matter to patients
  - Practical measures
- Coverage of a wide range
- Greater precision
- Comparability of scores
- Ease of interpretation
What Do We Need to Measure Health Outcomes in the 21st Century?

- Outcomes that matter to patients
- Practical measures
  - Coverage of a wide range
- Greater precision
- Comparability of scores
- Ease of interpretation
Problems with Short-Form Surveys

Measuring Too Low - Ceiling Effect
Some Thermometers Focus on a Very Narrow Range

Cooking Thermometer

130–190 °F
54–88 °C
A Promising Solution in 1999: Computerized Adaptive Testing (CAT) Software

Skewed 5-Item Headache Pain Measure

Dynamic 5-Item Headache Pain Measure

“Ceiling Effect”

Criterion Score

Criterion Score

r = 0.536
N = 1016
3 SD units

No Disability

r = 0.938
N = 1016

Improving the Physical Function “Ruler”

- **1980** Old “Ruler” > 75% @ Ceiling
- **1990** NEW “Ruler” > 30% @ Ceiling
- **2008** BETTER “Ruler” < 3% @ Ceiling

- **Vigorous Activities, Not limited**
- **Climbing several flights of stairs**
- **Walk one hundred yards**
- **Bathing or dressing, Limited a little**

**Norm**

**Mean = 50**
**SD = 10**
What Do We Need to Measure Health Outcomes in the 21st Century?

- Outcomes that matter to patients
- Practical measures
- Coverage of a wide range
- **Greater precision**
- Comparability of scores
- Ease of interpretation
Results were not interpretable
• No marks on the “ruler”
• Poor reproducibility
• No interpretation guidelines

Original Thermoscope

Visual Analogue Scale (VAS)
What Do We Need to Measure Health Outcomes in the 21st Century?

- Outcomes that matter to patients
- Practical measures
- Coverage of a wide range
- Greater precision
- **Comparability of scores**
- Ease of interpretation
**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

Note: Direction of scoring shown with arrows

<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>

Source: Ware, Bjorner & Kosinski, *Medical Care*, 2000

Theta ($\theta$) [Best Possible Estimate]
What Do We Need to Measure Health Outcomes in the 21st Century?

- Outcomes that matter to patients
- Practical measures
- Coverage of a wide range
- Greater precision
- Comparability of scores
- Ease of interpretation
Standardization

"Improvements in Short Form Measures of Health Status,"

*J. Clinical Epidemiology*, 2008
Interpreting Health Measures

Causes
- Diagnosis
- Disease severity
- Responders
- Treatments

Measures In Question

Gold Standard

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

Other Measures & Methods

What Do Changes in Health Mean?

• 50% reduction in physical disability
• 33% reduction in hospitalization
• Substantial increase in work productivity
• Reduced expenditures the following year
Advances & Solutions

- Improved psychometrics (Item response theory – IRT)
- Computerized adaptive testing (CAT) software
- The Internet (and other connectivity)
Precision Varies for “Static” and Dynamic Forms and Score Levels for Physical Function Measures

2nd Solution:
Assess Health Dynamically

CAT = Computerized Adaptive Testing
Dynamic Assessments Match Questions to Each Patient’s Level

- **Severe**
- **Moderate**
- **Mild**
First Question
Noisy Score Estimate (+/- 15)

First Item
Score = 62 +/- 15

Score Levels:
- Mild
  - 10 to 30
- Moderate
  - 30 to 50
- Severe
  - 50 to 80
Second Question: Standard Error Reduced by One Third

Second Item Score = 64 +/- 10
Third Question: Standard Error Cut in Half

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

Score = 62 ± 5

Severe
Moderate
Mild

Fourth Item
Score = 62
+/- 5
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 numerous ways
• Markedly reduced data collection costs
• Monitor data quality in real time
3rd Solution: The Internet
Internet Health Portals for Sampling, Data Capture and Reporting


Health Outcomes Research Using iPods/ iPhones
Conclusions

- Patient-reported outcomes (PROs) are very useful
- Standardization of concepts & metrics is enabling comparisons & interpretation across applications
- Increasing widespread use proves that more practical tools will be adopted
- Technological advances include: item response theory (IRT), computerized adaptive testing (CAT) and electronic data capture
- Data quality studies support use of improved measurement tools across diverse populations