Measurement of Outcomes Ron D. Hays

<u>Accelerating eXcellence In</u> translational <u>Science</u> (AXIS)

January 17, 2013 (2:00-3:00 pm)

1720 E. 120th Street, L.A., CA 90059

Conference Room 3071 (Hawkins)

Time Allocation

- Lecture
 - 40 minutes
- Q&A
 - 10 minutes
- Quiz
 - 10 minutes



U.S. Health Care Issues

• Access to care

 $- \sim 50$ million people without health insurance

- Costs of care
 - Expenditures ~ \$ 2.7 Trillion

• Effectiveness (quality) of care

How Do We Know If Care Is Effective?

- Effective care maximizes probability of desired health outcomes
 - Outcomes are markers of whether or not care is effective

Cost ↓

Effectiveness ↑

What Are Health Outcomes?

- Traditional clinical endpoints
 - Death, disease occurrence, other adverse events
 - Clinical measures/biological indicators
 - Blood pressure
 - Blood hemoglobin level



Patient-Reported Measures (PRMs)

- - Outcomes (PROs) Health-Related Quality of Life (HRQOL)
 - Satisfaction with care
- Mediators
 - Health behaviors (adherence)
- Health Care Process - Reports about care (e.g., communication)



Health-Related Quality of Life is:

How the person FEELs (well-being)

- Emotional well-being
- Pain
- Energy

What the person can DO (functioning)

- Self-care
- Role
- Social

Does your health now limit you in walking more than a mile?

(If so, how much?)

Yes, limited a lot Yes, limited a little No, not limited at all

How much of the time during the past 4 weeks have you been happy?

None of the time A little of the time Some of the time Most of the time All of the time





www.nihpromis.org

- Patient Reported Outcomes Measurement Information System (PROMIS)
- Funded by the National Institutes of Health
- System of reliable, valid, flexible, precise, and responsive HRQOL measures
- One domain captured is "anger"
 - Mood (irritability, frustration), negative social cognitions (interpersonal sensitivity, envy, disagreeableness) and efforts to control anger

Reliability (0-1)

- 0.70 or above for group comparisons
- 0.90 or above for individual assessment
- z-scores (mean = 0 and SD = 1):
 - Reliability = $1 SE^2$
 - So reliability = 0.90 when SE = 0.32
- T-scores (mean = 50 and SD = 10):
 - Reliability = $1 (SE/10)^2$
 - So reliability = 0.90 when SE = 3.2

In the past 7 days

- I was grouchy [1st question]
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always
- •Theta = 56.1 SE = 5.7 (rel. = 0.68)

I felt like I was ready to explode [2nd question]

- Never
- Rarely
- Sometimes
- Often
- Always

•Theta = 51.9 SE = 4.8 (rel. = 0.77)

- I felt angry [3rd question]
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always

• Theta = 50.5 SE = 3.9 (rel. = 0.85)

I felt angrier than I thought I should [4th question]

- Never
- Rarely
- Sometimes
- Often
- Always

•Theta = 48.8 SE = 3.6 (rel. = 0.87)

- I felt annoyed [5th question]
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always
- •Theta = 50.1 SE = 3.2 (rel. = 0.90)

I made myself angry about something just by thinking about it. [6th question]

- Never
- Rarely
- Sometimes
- Often
- Always

•Theta = 50.2 SE = 2.8 (rel = 0.92)

Theta, SEM, and 95% CI

- 56 and 6 (reliability = .68) W = 22
- 52 and 5 (reliability = .77) W = 19
- 50 and 4 (reliability = .85) W = 15
- 49 and 4 (reliability = .87) W = 14
- 50 and 3 (reliability = .90) W = 12
- 50 and <3 (reliability = .92) W = 11

Emotional Well-Being and Physical Functioning of 54 Patients at UCLA-Center for East-West Medicine



MS = multiple sclerois; ESRD = end-stage renal disease; GERD = gastroesophageal reflux disease.

Effect Size

(Follow-up – Baseline)/ SD_{baseline}

Cohen's Rule of Thumb:

- \checkmark ES = 0.20 Small
- \checkmark ES = 0.50 Medium
- \checkmark ES = 0.80 Large

Effect Sizes for Changes in SF-36 Scores

Effect Size



PFI = Physical Functioning; Role-P = Role-Physical; Pain = Bodily Pain; Gen H=General Health; Energy = Energy/Fatigue; Social = Social Functioning; Role-E = Role-Emotional; EWB = Emotional Well-being; PCS = Physical Component Summary; MCS = Mental Component Summary.

Significant Improvement in all but 1 of SF-36 Scales (Change is in T-score metric)

| | Change | t-test | prob. |
|-------------|--------|--------|---------|
| PF-10 | 1.7 | 2.38 | .0208 |
| RP-4 | 4.1 | 3.81 | .0004 |
| BP-2 | 3.6 | 2.59 | .0125 |
| GH-5 | 2.4 | 2.86 | .0061 |
| EN-4 | 5.1 | 4.33 | .0001 |
| SF-2 | 4.7 | 3.51 | .0009 |
| RE-3 | 1.5 | 0.96 | .3400 ← |
| EWB-5 | 4.3 | 3.20 | .0023 |
| PCS | 2.8 | 3.23 | .0021 |
| MCS | 3.9 | 2.82 | .0067 |

Defining a Responder: Reliable Change Index (RCI)



 $SEM = SD_{hl} \times \sqrt{1 - r_{xx}}$

Note: SD_{bl} = standard deviation at baseline r_{xx} = reliability

Amount of Change in Observed Score Needed for Significant Individual Change

| Scale | RCI | Effect size | Reliability |
|-------|------|-------------|--------------------|
| PF-10 | 8.4 | 0.67 | 0.94 |
| RP-4 | 8.4 | 0.72 | 0.93 |
| BP-2 | 10.4 | 1.01 | 0.87 |
| GH-5 | 13.0 | 1.13 | 0.83 |
| EN-4 | 12.8 | 1.33 | 0.77 |
| SF-2 | 13.8 | 1.07 | 0.85 |
| RE-3 | 9.7 | 0.71 | 0.94 |
| EWB-5 | 13.4 | 1.26 | 0.79 |
| PCS | 7.1 | 0.62 | 0.94* |
| MCS | 9.7 | 0.73 | 0.93* |

* Mosier's formula for weighted combination of scales; coefficient alpha for others.

7-31% of People in Sample Improve Significantly

| | % Improving | % Declining | Difference |
|-------------|-------------|-------------|------------|
| PF-10 | 13% | 2% | + 11% |
| RP-4 | 31% | 2% | + 29% |
| BP-2 | 22% | 7% | + 15% |
| GH-5 | 7% | 0% | + 7% |
| EN-4 | 9% | 2% | + 7% |
| SF-2 | 17% | 4% | + 13% |
| RE-3 | 15% | 15% | 0% |
| EWB-5 | 19% | 4% | + 15% |
| PCS | 24% | 7% | + 17% |
| MCS | 22% | 11% | + 11% |

Preference-Based HRQOL Measure (0 = dead, 1 = perfect health, -- = worse than being dead)



Questions?



Contact Information: <u>drhays@ucla.edu</u> (310-794-2294) Powerpoint file available for downloading at: <u>http://gim.med.ucla.edu/FacultyPages/Hays/</u>