Psychometric Modeling and Calibration

Ron D. Hays, Ph.D. September 11, 2006

PROMIS development process session 10:45am-12:30 pm

#### **PROMIS Domains**

**Physical functioning (Hays/Bjorner)** 

Pain (Revicki)

**Fatigue (Lai)** 

**Emotional distress (Choi/Reise)** 

Social/role participation (Bode/Hahn)

#### **Datasets**

**Cancer Item Banks (Northwestern)** 

Digitalis Investigation Group Study--randomized doubleblind placebo-controlled trial evaluating effect of digoxin on mortality in 581 patients with heart failure and sinus rhythm.

IMMPACT--internet-based survey of individuals with chronic pain from the American Chronic Pain Association website

Medical Outcomes Study--observational study of persons with hypertension, diabetes, heart disease, and/or depression in Boston, Chicago, and Los Angeles

WHOQOL-100 data (n = 442 from U.S. field center)

# **Types of Analyses**

Classical Test Theory Statistics

IRT Model Assumptions

Model Fit

Differential Item Functioning

Item Calibration

#### **Classical Test Theory Statistics**

Out of range

Item frequencies and distributions

Inter-item correlations

Item-scale correlations

Internal consistency reliability

#### **IRT Model Assumptions**

(Uni)dimensionality

Local independence

Monotonicity

# **Sufficient Unidimensionality**

Confirmatory factor models

One factor

Bifactor (general and group factors)

# Local Independence

 After controlling for dominant factor(s), item pairs should not be associated.

Look at residual correlations (> 0.20)

# Monotonicity

- Probability of selecting a response category indicative of better health should increase as underlying health increases.
- Item response function graphs with

   y-axis: proportion positive for item step
   x-axis: raw scale score minus item score



10 1/23/18

# Category Response Curves for Samejima's Graded Response Model



# **Model Fit**

- Compare observed and expected response frequencies by item and response category
- Items that do not fit and less discriminating items identified and reviewed by content experts

# **Differential Item Functioning**

Uniform DIF

Threshold parameter

Non-uniform DIF

Discrimination parameter

Gender, race/ethnicity, age, disease

#### **Item Calibration**

- Item parameters (threshold, discrimination)
- Mean differences for studied disease groups

Example of Lessons Learned in Secondary Analyses

**Emotional distress** 

Cannot be adequately modeled as a unidimensional construct.

Limited representation of positive end of construct

Several items having some response options that provide little information.

#### **Documentation**

eRoom

Public website: http://www.nihpromis.org/

Peer-reviewed manuscripts, e.g.:

Hays, R. D. et al. (submitted). Item response theory analyses of physical functioning items in the Medical Outcomes Study.

Reeve, B. B. (submitted). Psychometric evaluation and calibration of health-related quality of life items banks: Plans for the Patient-Reported Outcome Measurement Information System (PROMIS)

#### **Presentations:**



Tomorrow 4:15-6:15

# **Questions?**



#### **Datasets Subjected to Psychometric Analysis**

Cancer Fatigue: Cancer Item Banking Project at NWU Cancer Pain: Cancer Item Banking Project at NWU Cancer Social: Cancer Item Banking Project at NWU **CSSCD:** Cooperative Study of Sickle Cell Disease (pediatric) **CHC:** Chronic Hepatitis C Study **CHS:** Cardiovascular Health Study **DIG:** Digitalis Investigation Group Quality of Life Sub-study **IMMPACT: Multiple Pain Projects MOS:** Medical Outcomes Study **NGHS:** National Growth and Health Study (peds.) **Q-Score:** Cancer Quality of Life Project at NWU WHOQOL: World Health Organization Quality of Life Project

#### **Datasets Subjected to Psychometric Analysis**

