

Considerations in Comparing Groups of People with PROs

Ron D. Hays, Ph.D.

UCLA Department of Medicine

May 6, 2008, 3:45-5:00pm

ISPOR, Toronto, Canada



Measures taken from the patient

140/90

98.6° F or 37.0° C

36-24-36



Measures given by the patient

“My health is excellent.”

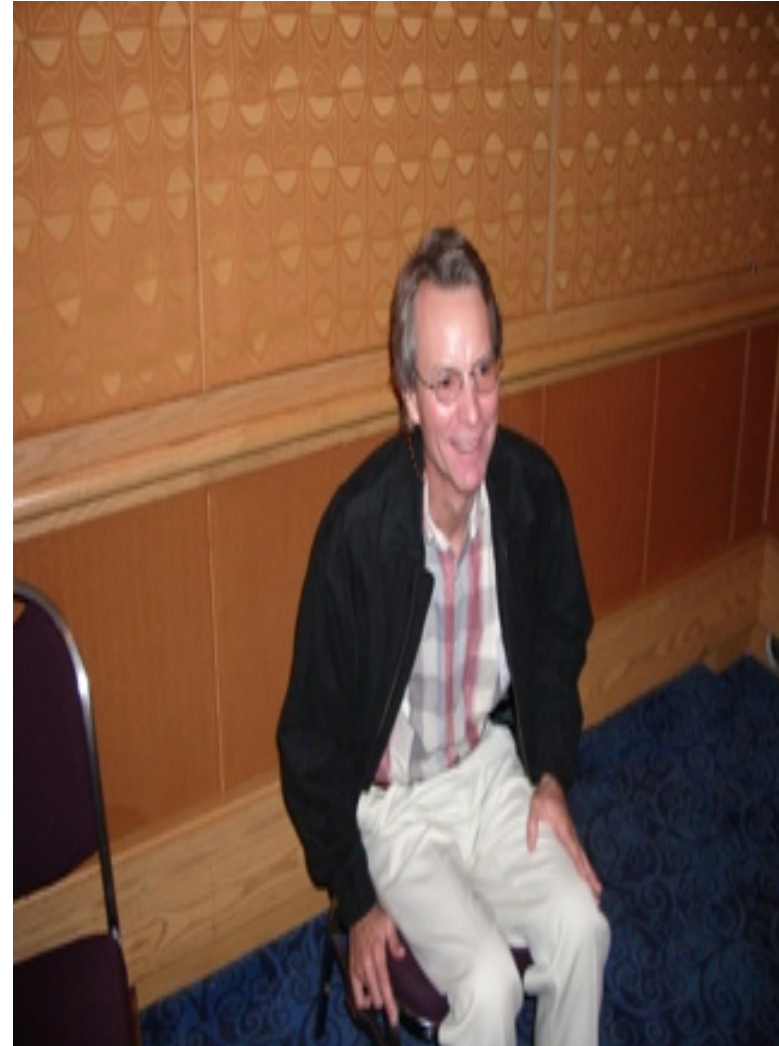
“I get along with my wife.”

“I have a lot of energy.”

“I am good at my job.”

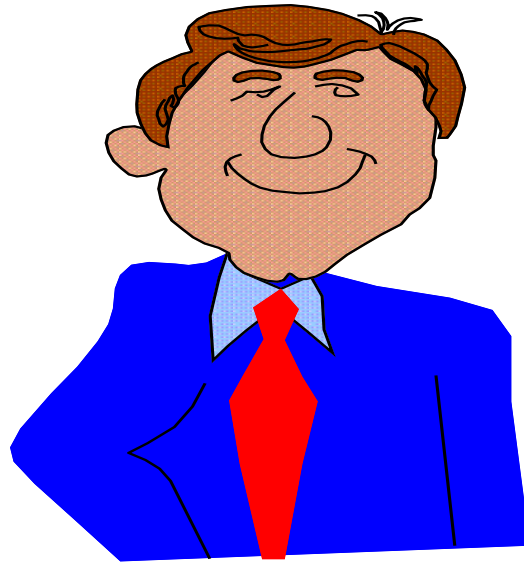
“I can walk a block.”

“My vision sucks.”



Health-Related Quality of Life

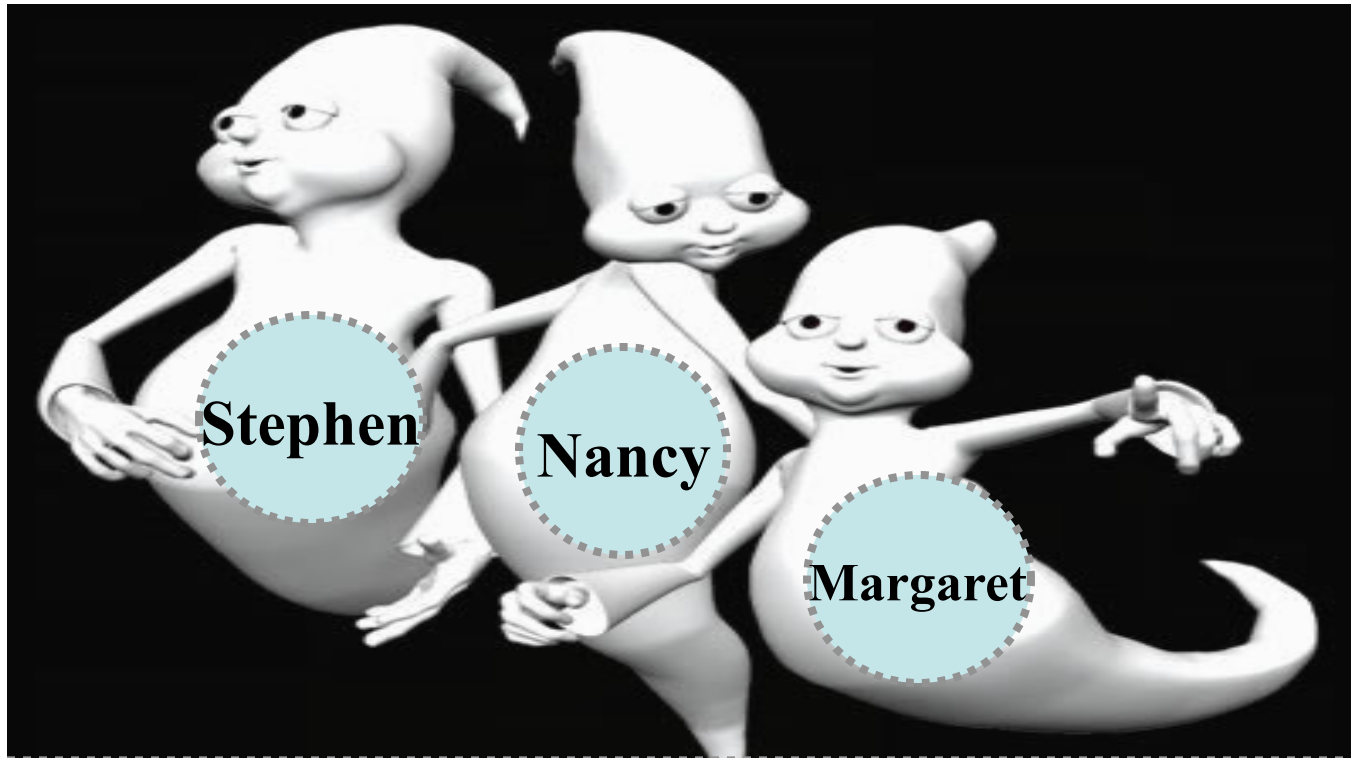
What they are able to do



Thanks for asking, doc

And how they feel about their life

Skepticism Can be Healthy



2/3 of Canadians believe in angels and 1/2 in spirits and ghosts.

SKEPTIC

“The public believes x-rays and biopsies give clear yes-or-no answers, but that is not the case. They depend on human perception, pattern-recognition, and interpretation.

Pathologists and radiologists often disagree with each other and even with themselves (in repeat readings). A radiologist who has recently missed a breast cancer is likely to over-read future mammograms in compensation.”

<http://www.skeptic.com/eskeptic/07-10-24.html>

SPECIAL ARTICLE

Zoom In

Precision of Health-Related Quality-of-Life Data Compared With Other Clinical Measures

ELIZABETH A. HAHN, MA; DAVID CELLA, PhD; OLIVIER CHASSANY, MD, PhD; DIANE L. FAIRCLOUGH, DRPH;
GILBERT Y. WONG, MD; RON D. HAYS, PhD; AND THE CLINICAL SIGNIFICANCE CONSENSUS MEETING GROUP

To many clinicians, the assessment of health-related quality of life (HRQL) seems more art than science. This belief is due in part to the lack of formal training available to clinicians regarding HRQL measurement and interpretation. When HRQL is used systematically, it has been shown to improve patient-physician communication, clinical decision making, and satisfaction with care. Nevertheless, clinicians rarely use formal HRQL data in their practices. One major reason is unfamiliarity with the interpretation and potential utility of the data. This unfamiliarity causes a lack of appreciation for the reliability of data generated by formal HRQL assessment and a tendency to regard HRQL data as having insufficient precision for individual use. This article discusses HRQL in the larger context of health indicators and health outcome measurement and is targeted to the practicing clinician who has not had the opportunity to understand and use HRQL data. The concept and measurement of reliability are explained and applied to HRQL and common clinical measures simultaneously, and these results are compared with one another. By offering a juxtaposition of common medical measurements and their associated error with HRQL measurement error, we note that HRQL instruments are comparable with commonly used clinical data. We further discuss the necessary requirements for clinicians to adopt formal, routine HRQL assessment into their practices.

Mayo Clin Proc. 2007;82(10):1244-1254

have demonstrated a positive impact on patient satisfaction or HRQL.^{10,13,17-20}

Assessment of HRQL has been successfully used to change and influence patient and physician communication, resulting in improved patient satisfaction in a community practice setting.²¹ The mechanisms by which routine assessment of HRQL might improve clinical practice include (1) aiding detection of physical or psychosocial problems that otherwise might be overlooked, (2) monitoring disease and treatment, (3) allowing precisely timed alterations in therapeutic plans, (4) facilitating patient-physician communication, and (5) improving the delivery of care.^{18,19,22-29} It is also possible to routinely use HRQL instruments in clinical practice to evaluate the efficacy of interventions designed to prevent or treat common problems experienced by patients.³⁰ Several critical elements for the success of routine HRQL assessments have been identified.^{15,26,31,32} The first is the availability of an acceptable set of measures from which to choose. These HRQL

Behavioral Risk Factor Surveillance System (BRFSS)

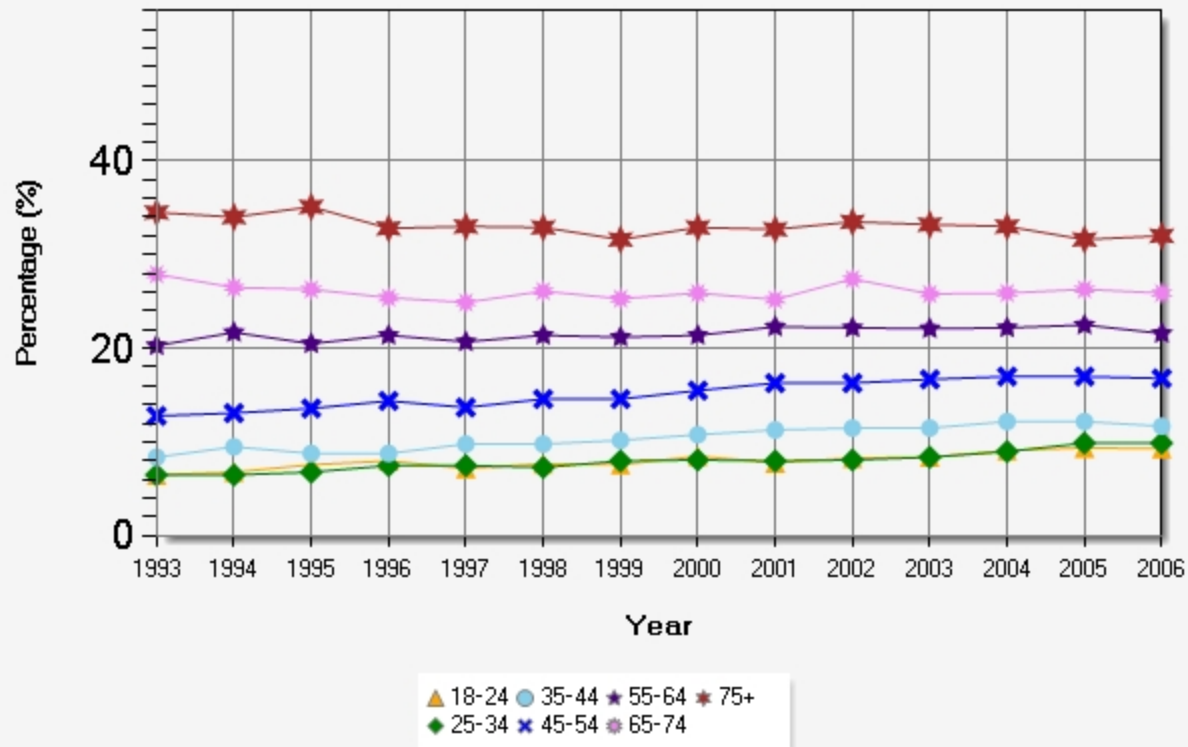
Nationwide telephone interview (random
digit dialing) of U.S. adults

*“Would you say in general your health is
excellent, very good, good, fair or poor?”*

Percent fair or poor health about **16%**

Percentage with fair or poor self-rated health

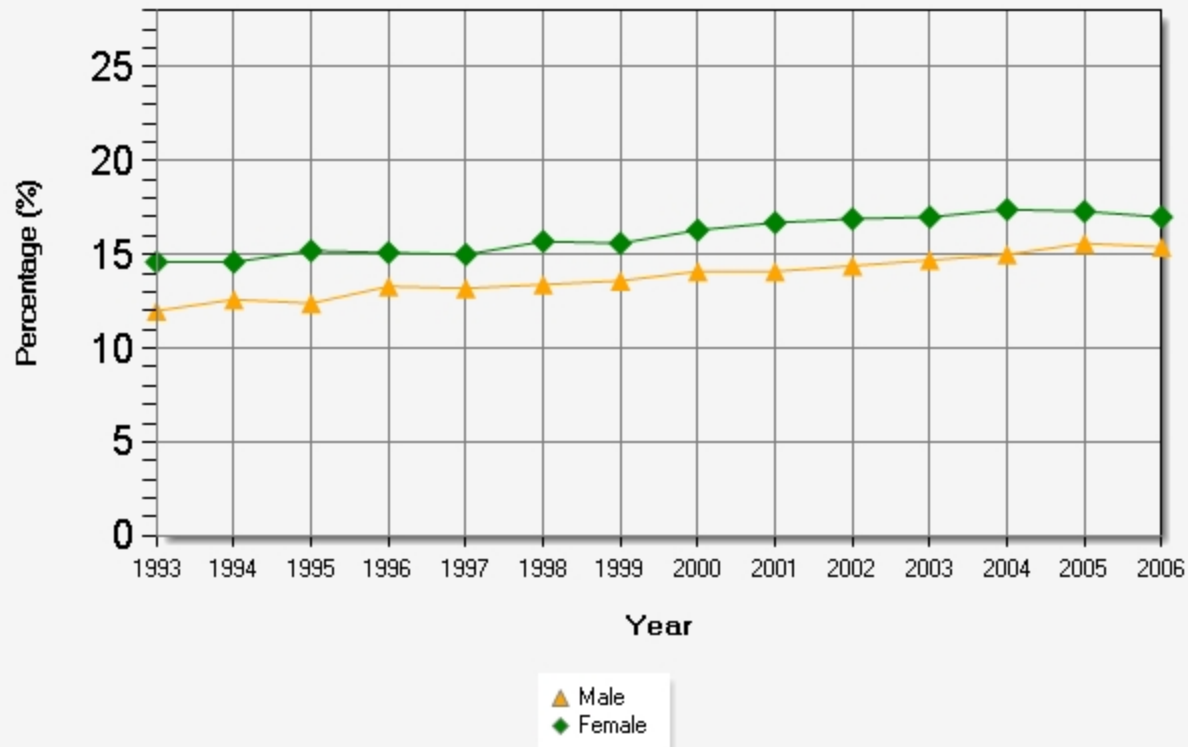
Nationwide trend: Age Group



Greater % of *fair* or *poor* health reported
by older adults (33% for 75+ vs. 9% for 18-24)

Percentage with fair or poor self-rated health

Nationwide trend: Gender



**Greater % of *fair* or *poor* health reported
by females (17%) vs. males (15%)**

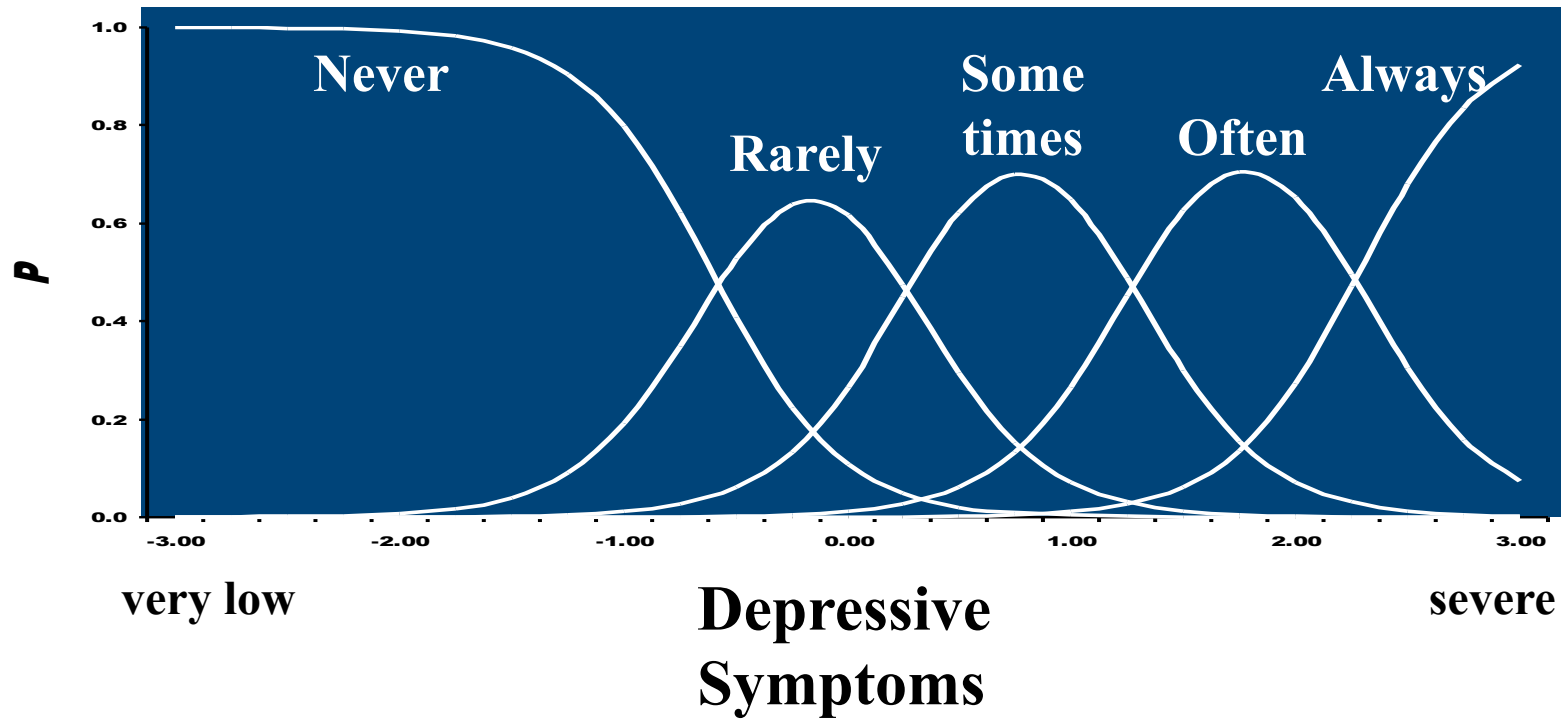
Don't Assume Equivalence

Evaluate by

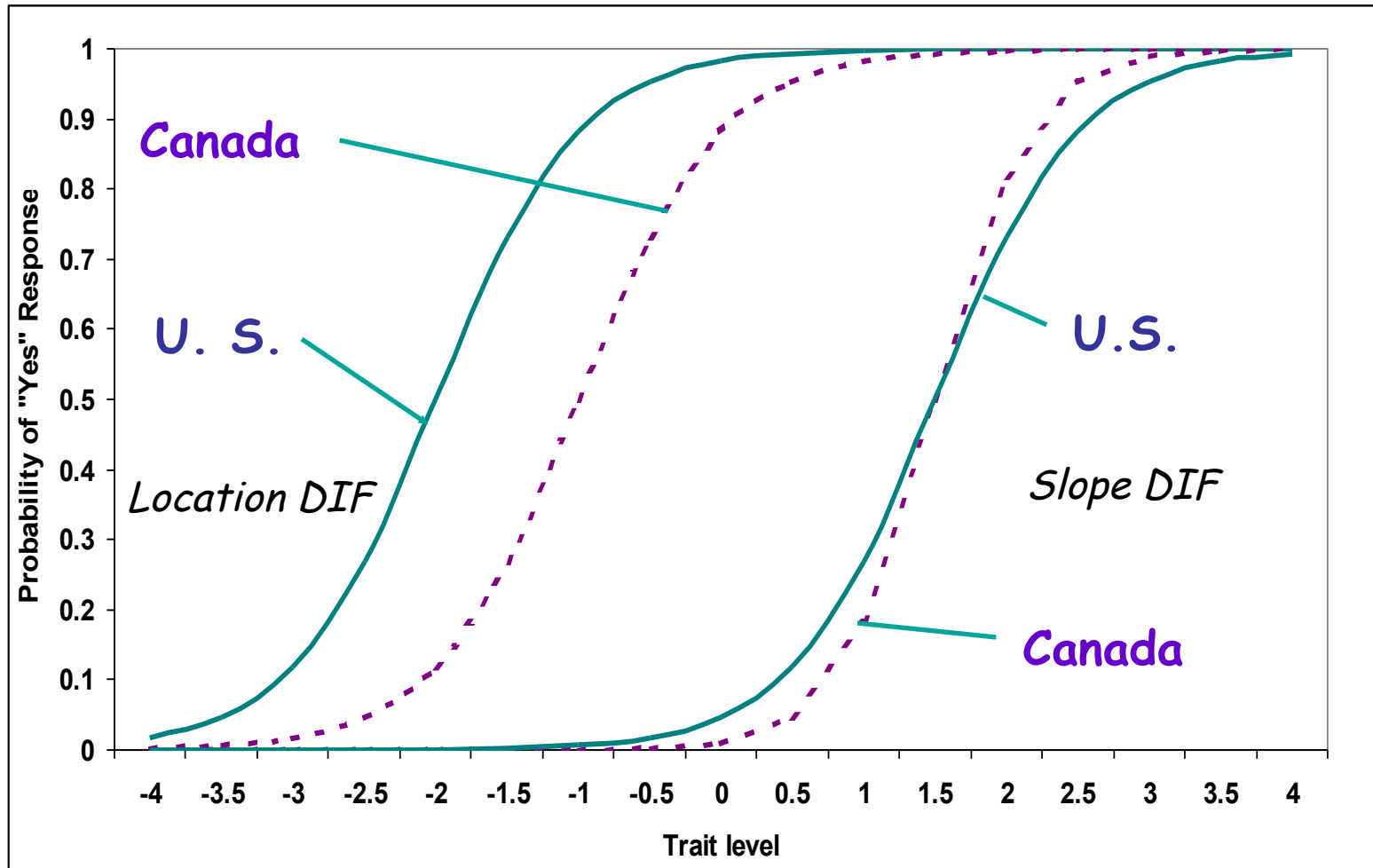
- Individual characteristics (age, gender, race, language, etc.)
- Site (including country)
- Administration effects
 - Order
 - Time
 - Mode
 - Form

Item Response Theory (IRT)

Category Response Curves



Differential Item Functioning (2-Parameter Model)



**Items from
Instrument
A**

**Items from
Instrument
B**

**Items from
Instrument
C**

**New
Items**

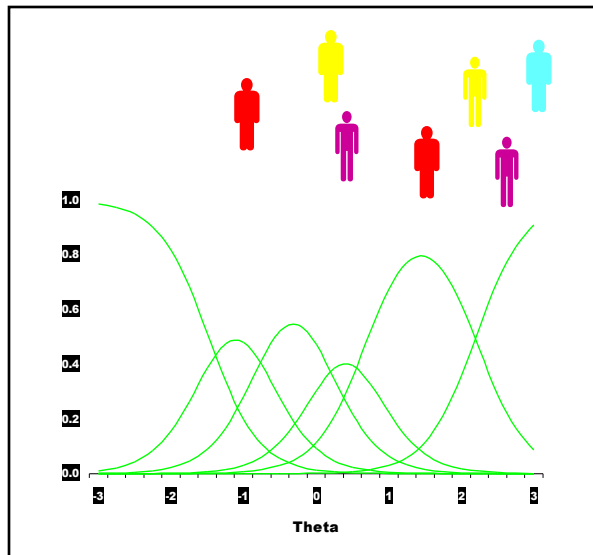
Item Pool

**Content Expert
Review**

**Focus
Groups**

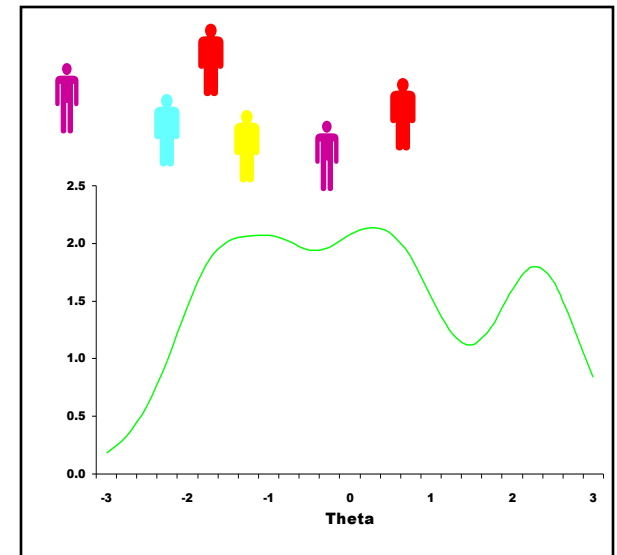
**Cognitive
Testing**

**Secondary
Data Analysis**



**Questionnaire
administered to large
representative sample**

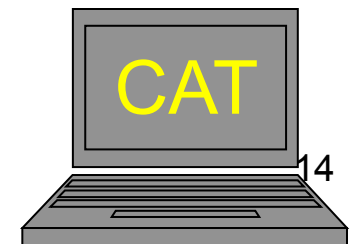
**Psycho-
metric
Testing**



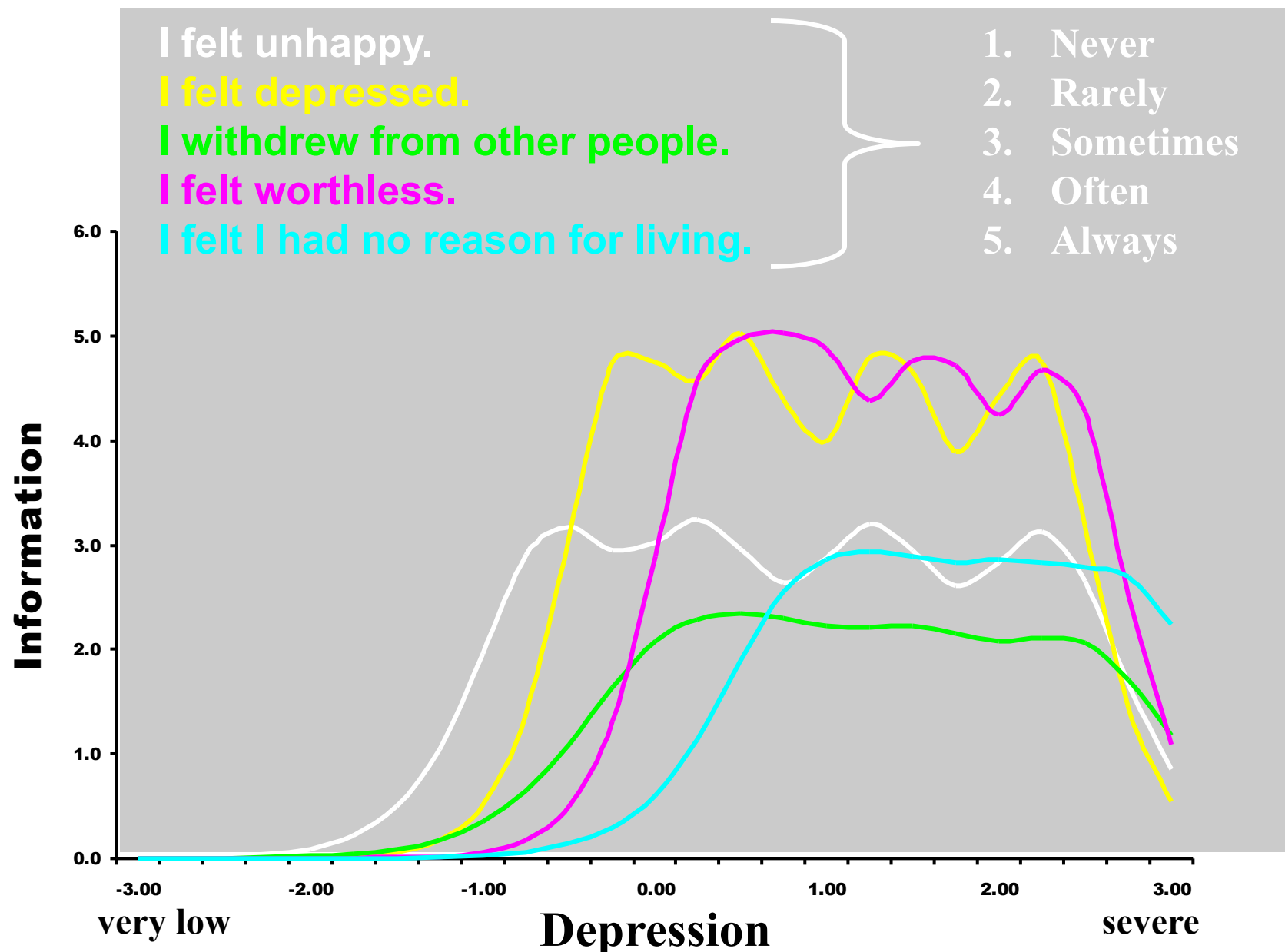
Item Bank

**(IRT-calibrated items reviewed for
reliability, validity, and sensitivity)**

**Short Form
Instruments**



Item Response Theory (IRT): Item Information Functions



Computerized Adaptive Testing (CAT)

- Estimate person's score (e.g., depressive symptoms) iteratively
- Administer most informative item
- Stop when desired level of precision is reached.

Respondent (n = 752)

Characteristics

Responded to the 28 items in the PROMIS depressive symptoms item bank

- Mean age = 51 (18-93 range)
- 52% female
- 78% white, 10% Hispanic, 10% black
- 22% high school graduate or less

Form A Items (Least to Most Severe)

#	<i>In the past 7 days...</i>
2	I felt disappointed in myself.
4	I felt discouraged about the future.
10	I found that things in my life were overwhelming.
18	I felt like a failure.
21	I felt that I had nothing to look forward to.
24	I felt that nothing could cheer me up.
26	I felt worthless.
28	I felt I had no reason for living.

Response Options: *Never, Rarely, Sometimes, Often, Always*

Form B Items (Least to Most Severe)

#	<i>In the past 7 days...</i>
1	I felt sad.
3	I felt unhappy.
9	I felt depressed.
17	I felt that nothing was interesting.
19	I felt that my life was empty.
23	I felt helpless.
25	I felt hopeless.
27	I felt I wanted to give up on everything.

Response Options: *Never, Rarely, Sometimes, Often, Always*

Mean, Minimum and Maximum Scores

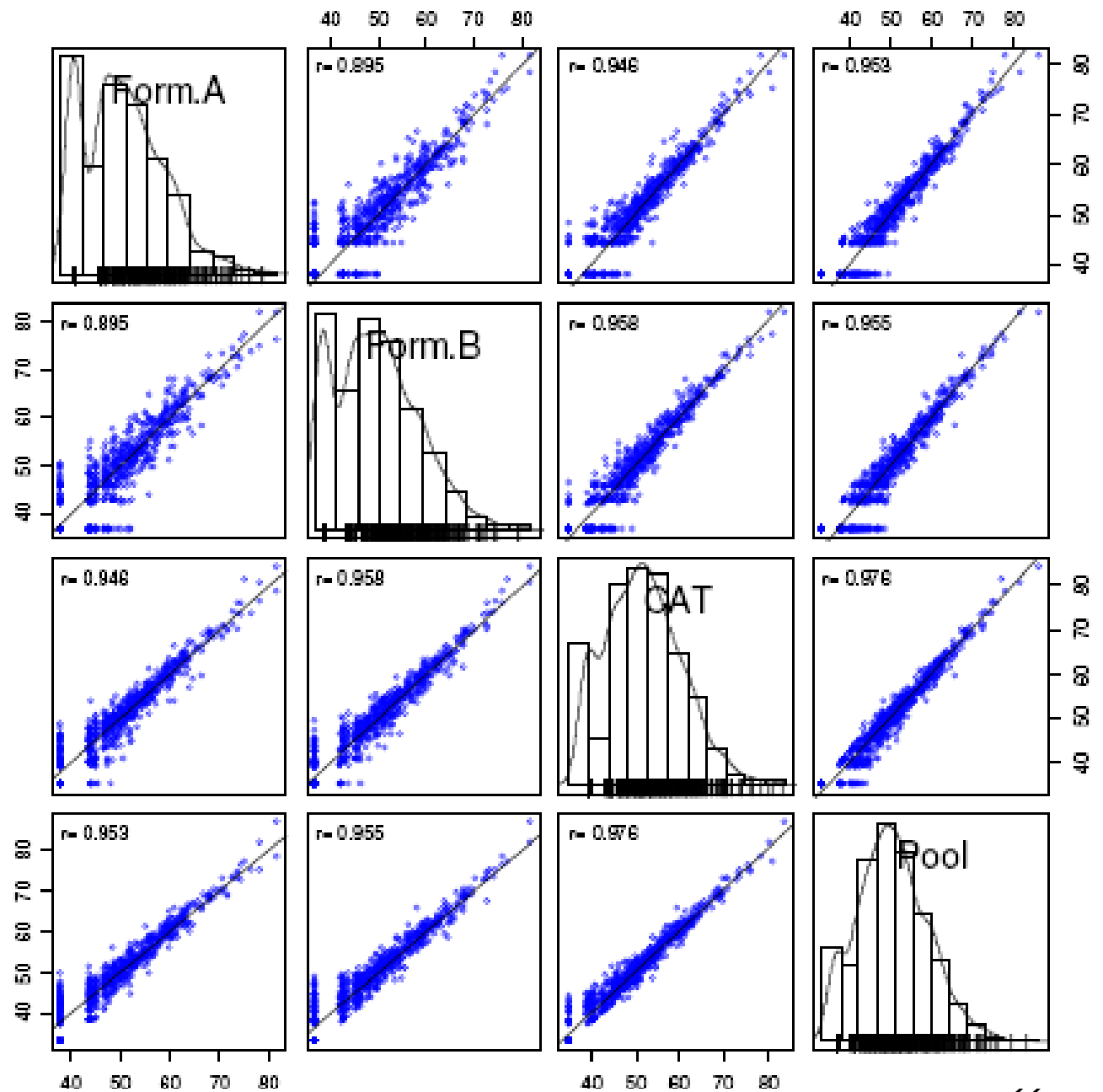
	Form A (8 items)	Form B (8 items)	CAT (8 items)	Full Bank (28 items)
Mean	49	49	49	49
Minimum	38	37	35	33
Maximum	82	82	84	86

752 respondents in PROMIS

Correlations Among 8-Item Forms and 28-Item Bank

	Form A	Form B	CAT
Form A	1.00		
Form B	0.89	1.00	
CAT	0.95	0.96	1.00
Bank	0.95	0.96	0.98

Scale Score Comparisons: Short Forms, CAT (8 items) and Full-Bank (28 Items)



Person Fit Can Also Be Evaluated

Z_L has expected value of zero, with variance of one (if person responds according to the estimated IRT model). Large negative Z_L values (≤ -2.0) indicate misfit.

Summary

- PROs (HRQOL) are as reliable as other measures used to assess patient health.
- Equivalence of PROs for different groups needs to be demonstrated.
- IRT provides strong empirical basis for evaluating equivalence



Thank you.

Ron D. Hays, Ph.D., UCLA Department of Medicine
911 Broxton Avenue, Room 110
Los Angeles, Ca 90095-1736 (drhays@ucla.edu)

**Preparation of this presentation was supported
in part by the National Institutes of Health
through the NIH Roadmap for Medical Research
Grant (AG015815), PROMIS Project.**

