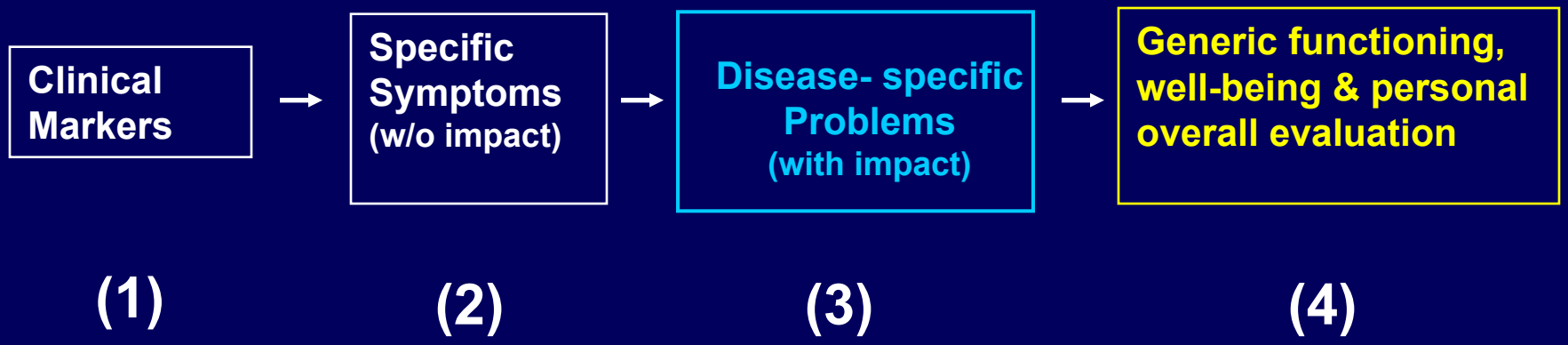
Future of Outcomes Measurement: Item Banking, Short Forms & CAT

John E. Ware, Jr., PhD

QualityMetric Incorporated, Lincoln, RI
Health Assessment Lab, Waltham, MA
Tufts University School of Medicine, Boston, MA

Advances in Health Outcomes Measurement
Co-sponsored by NCI, NIH, DHHS & DIA
Hyatt Regency, Bethesda, MD
June 23-25, 2004

Continuum of Disease-Specific and Generic Health Measures



Adapted from: Wilson and Cleary, *JAMA*, 1995
Ware, *Annual Rev. Pub. Health*, 1995

What is Health - Related Quality of Life (HRQOL)?

Ankle Brachial Index

Clinical Markers

→

Specific Symptom

Frequency of Pain

Specific Symptoms (w/o impact)

→

Impact-Specific

How much did your pain limit your usual activities or enjoyment of everyday life?
- Not at all
- A little
- Moderately
- Quite a lot
- Extremely

Disease- specific Problems (with impact)

→

Impact-Generic

SF-36 Health Survey Scales, Summary Measures

Generic functioning, well-being & personal overall evaluation
- HR-QOL profile
- HR-QOL summaries
- HR-QOL index

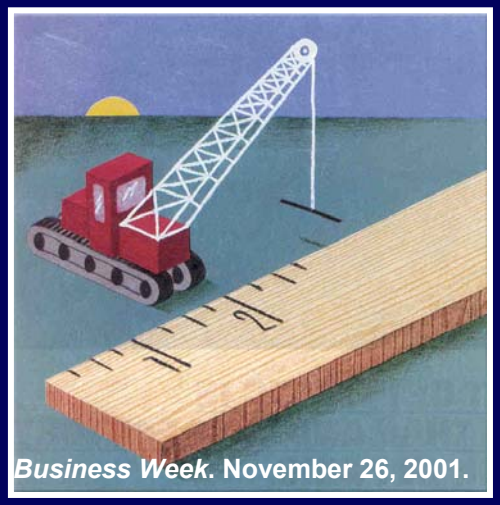
Health-Related QOL (HRQOL)

(1) (2) (3) (4)

References: Money et al., *J. Vasc Surg* 1998, 27(2):267-74
Hiatt WR, *New Eng J Med, 2001, 344(21), 160801621*
Regensteiner et al, *J Amer Geriat Soc,* 2002, 50, 1939-1946

What do we want from IRT and CAT?

- More practical tools
- More precise scores
- Measurement over a wider range
- Lower costs of data collection
- Greater comparability
- Results in real time



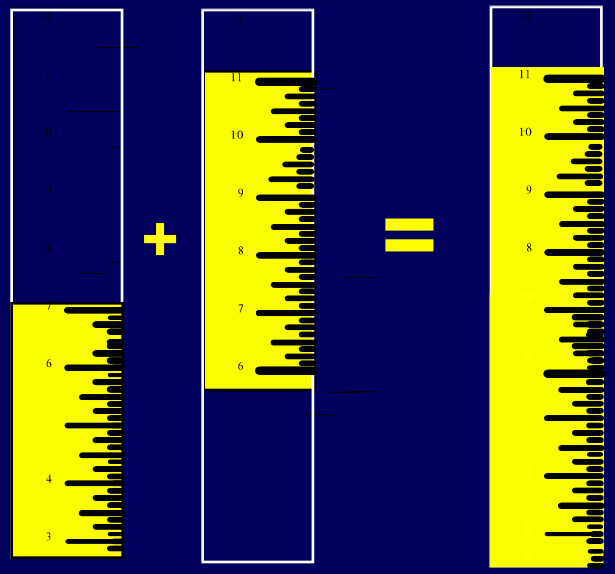
Business Week. November 26, 2001.

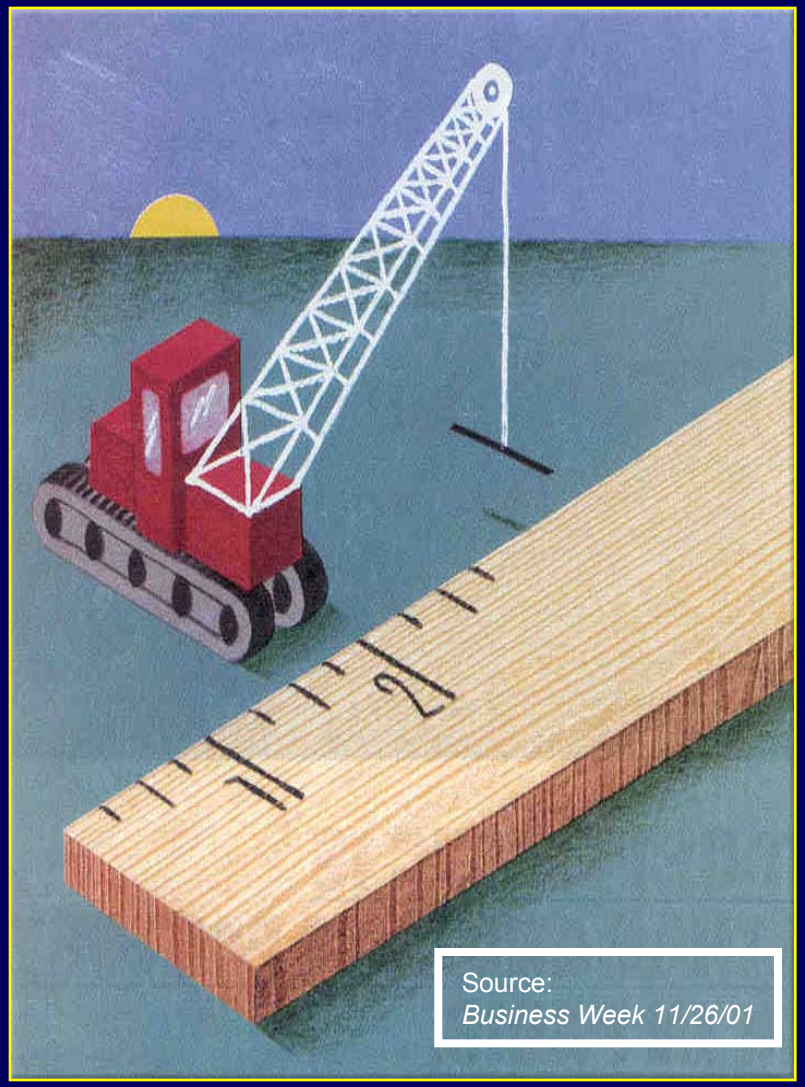
Solutions

- Improved psychometrics
- Computerized adaptive testing (CAT) software
- The Internet

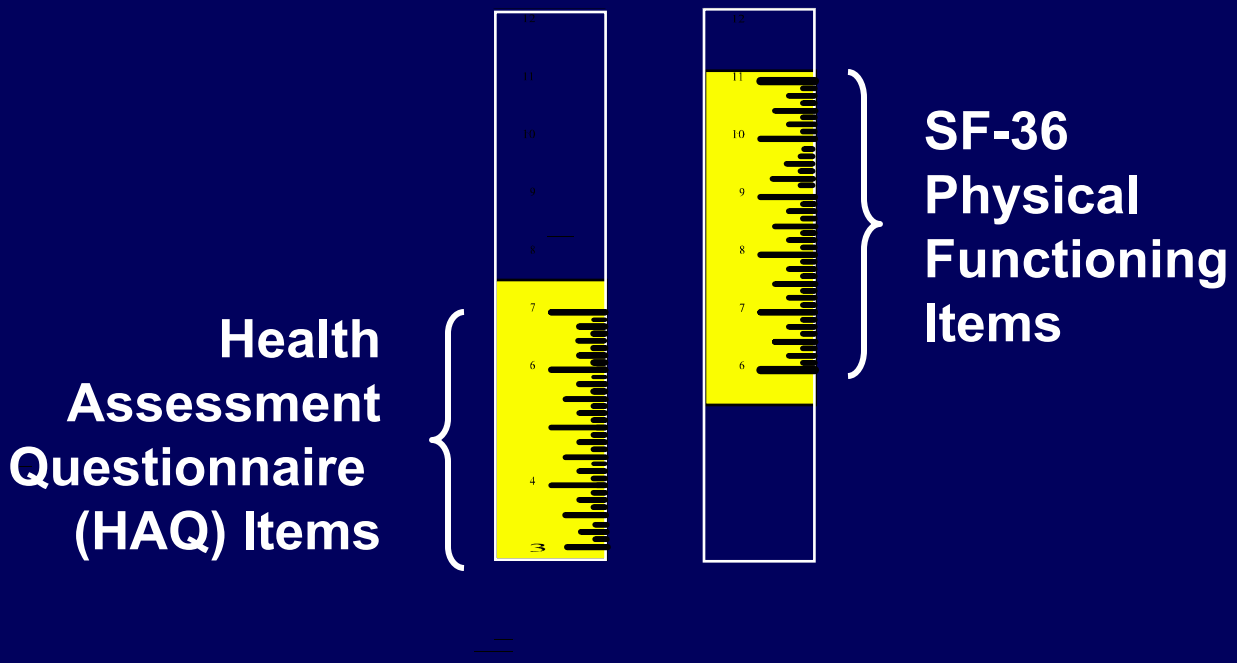
First, Construct Better Metrics

Item “Pools”
Item “Banking”
Cross Calibration



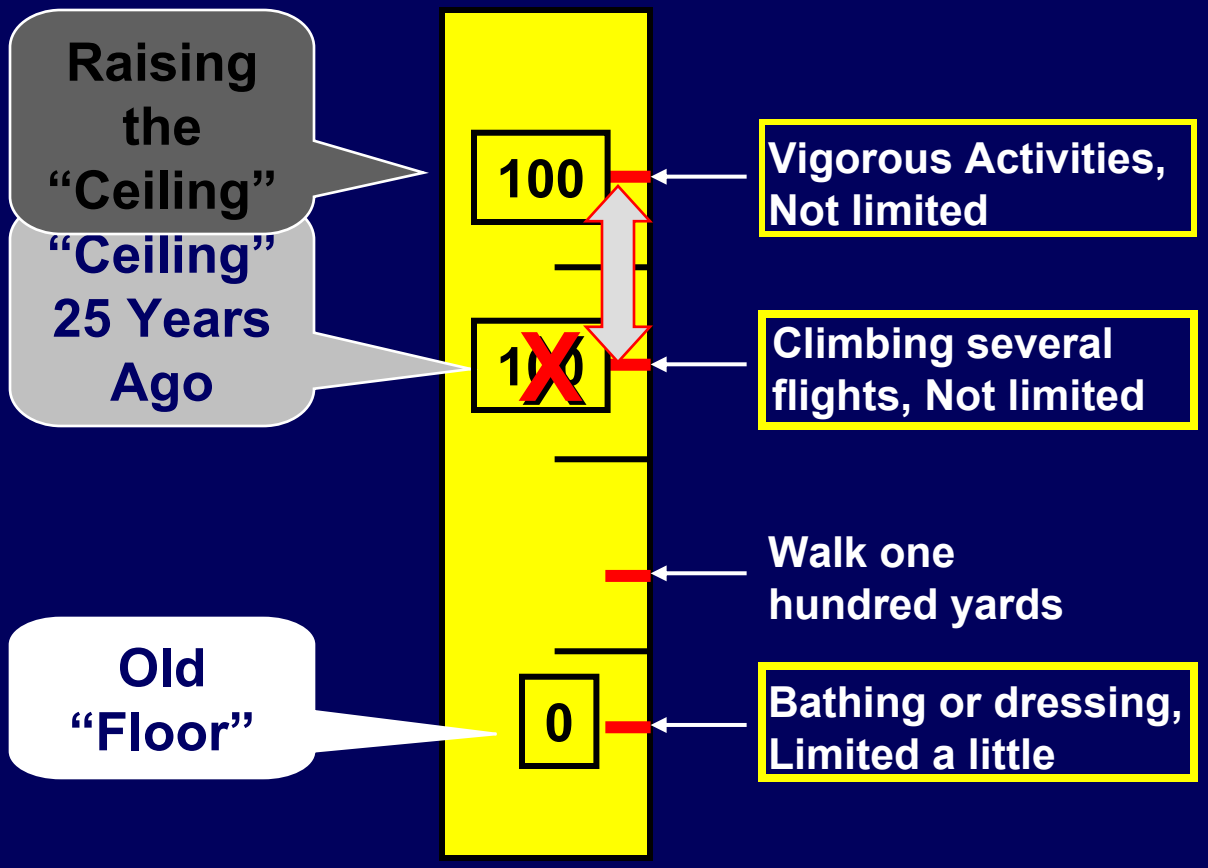


Combining HAQ & SF-36 Items Improves Measurement of Physical Functioning



Source: Martin, Kosinski, Bjorner et al., *Value in Health*, 2004

Original 0-100 Scoring Must Be Replaced



Physical Functioning

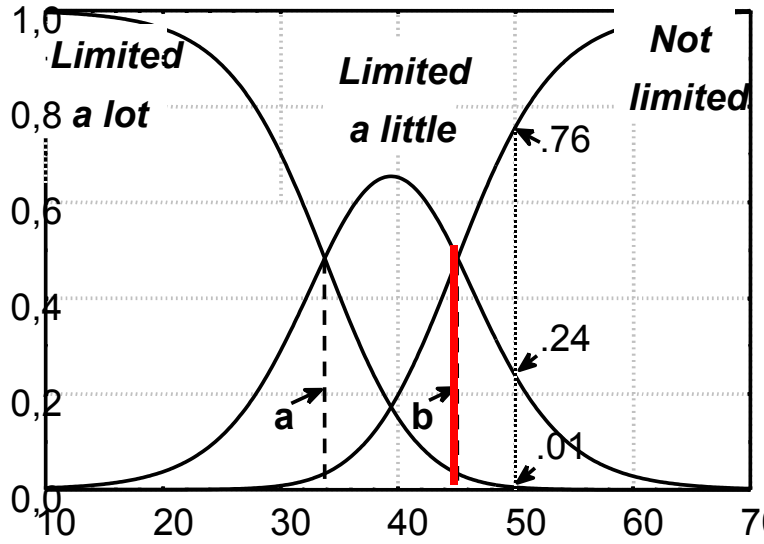
Original 0-100 Scoring Must Be Replaced

Physical Functioning

New "Ruler"
< 7 %

Old "Ruler"
> 30%
@ Ceiling

New F
Metr
S



100

50

0

60

40

30

20

Vigorous Acti

Norm

Climbing several flights, Not limited

Walk one hundred yards

Bathing or dressing, Limited a little

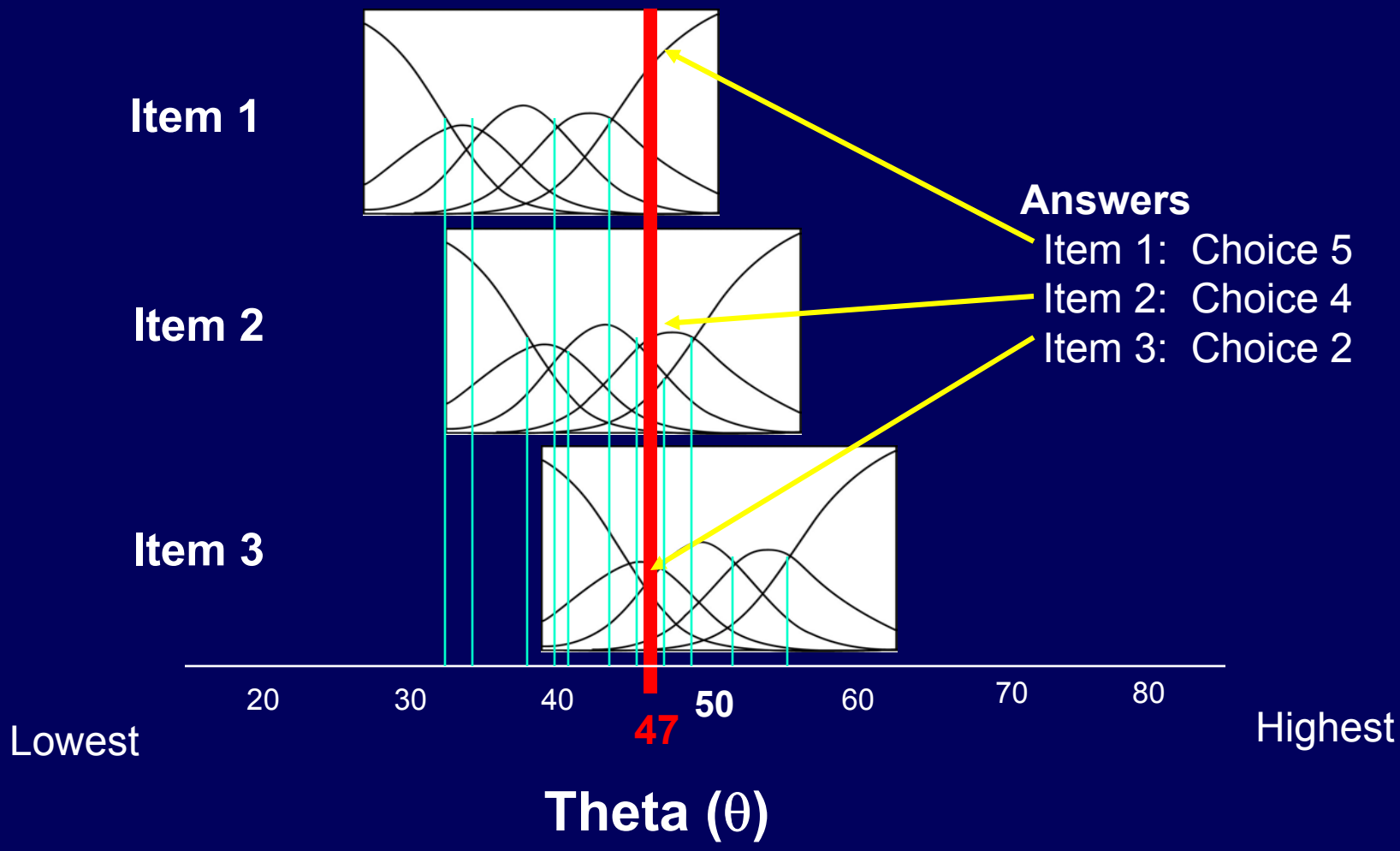
Mean = 50
SD = 10

Cross-Calibration Makes Scores Compar Interpretable

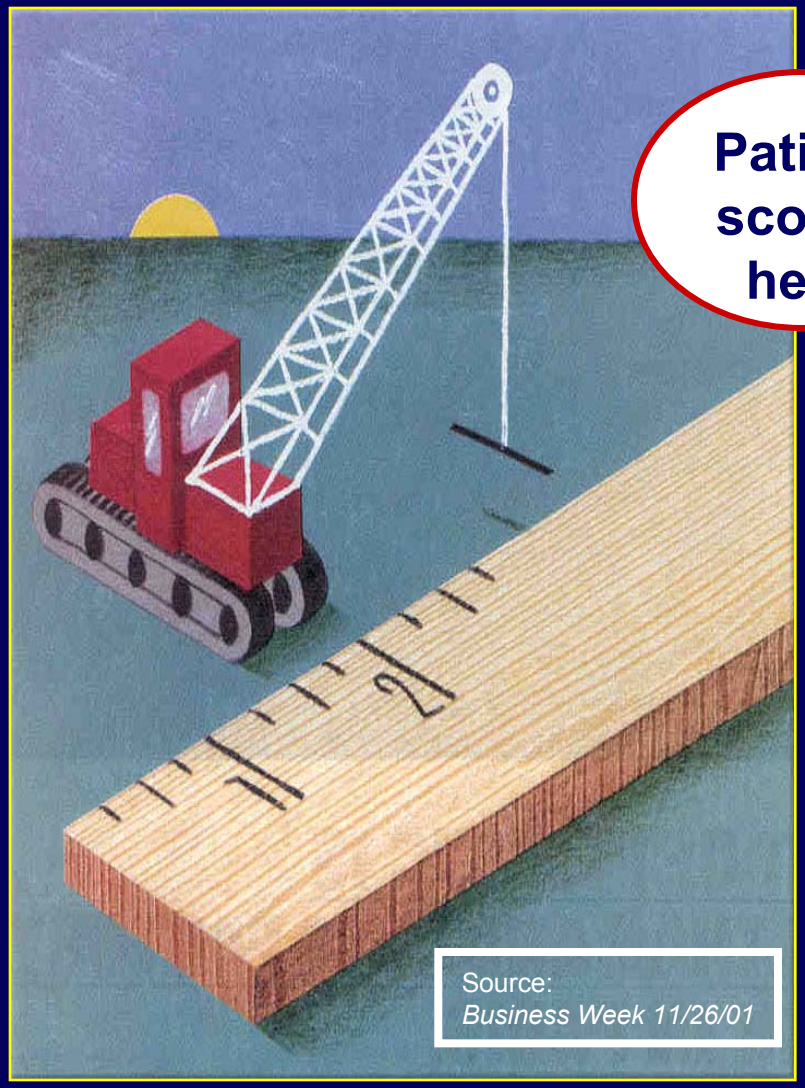
	Theta (θ) [Best Possible Estimate]					
Scales	**20**	**30**	**40**	**50**	**60**	**70**
HDI ↑	16	43	73	91	98	100
HIMQ ↓	74	53	31	17	8	2
MIDAS ↓	58	28	5	1	0	0
MSQ ↑	31	53	79	92	96	99
DYNHA-5 (+)	23	32	41	51	58	66

Note: Direction of scoring in parenthesis Source: Ware, Bjorner & Kosinski *Medical Care*, 2000

Response To Each Item is Predictable From Theta (θ)

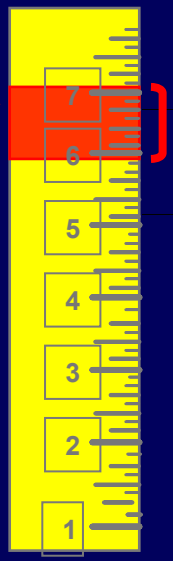


Second, Assess Health Dynamically



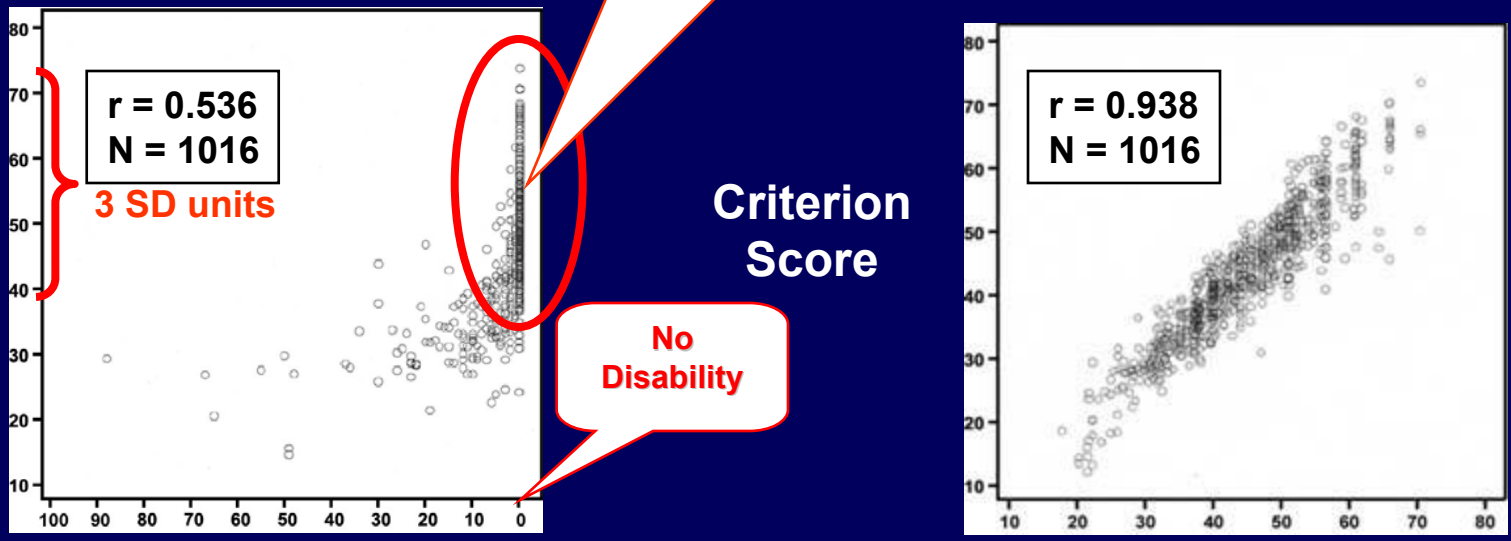
Patient scores here

Computerized Dynamic Health Assessment



Physical Functioning (PF)

Scatterplots Show That Some Short Forms Don't Measure Higher Levels of Health

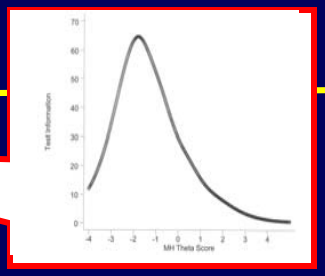


Ware JE, Jr, et al. *Med Care*. 2000;38:1173-82.

Logic of Computerized Adaptive Testing (CAT)

1. Begin with initial score estimate

2. Select & present optimal scale item



3. Score response

4. Re- estimate health score and confidence interval

5. Is stopping rule satisfied

No

Yes

6. End scale assessment

7. End of battery?

No

8. Administer next scale

Yes

9. Stop

Source: Adapted from Wainer et al. (2000)

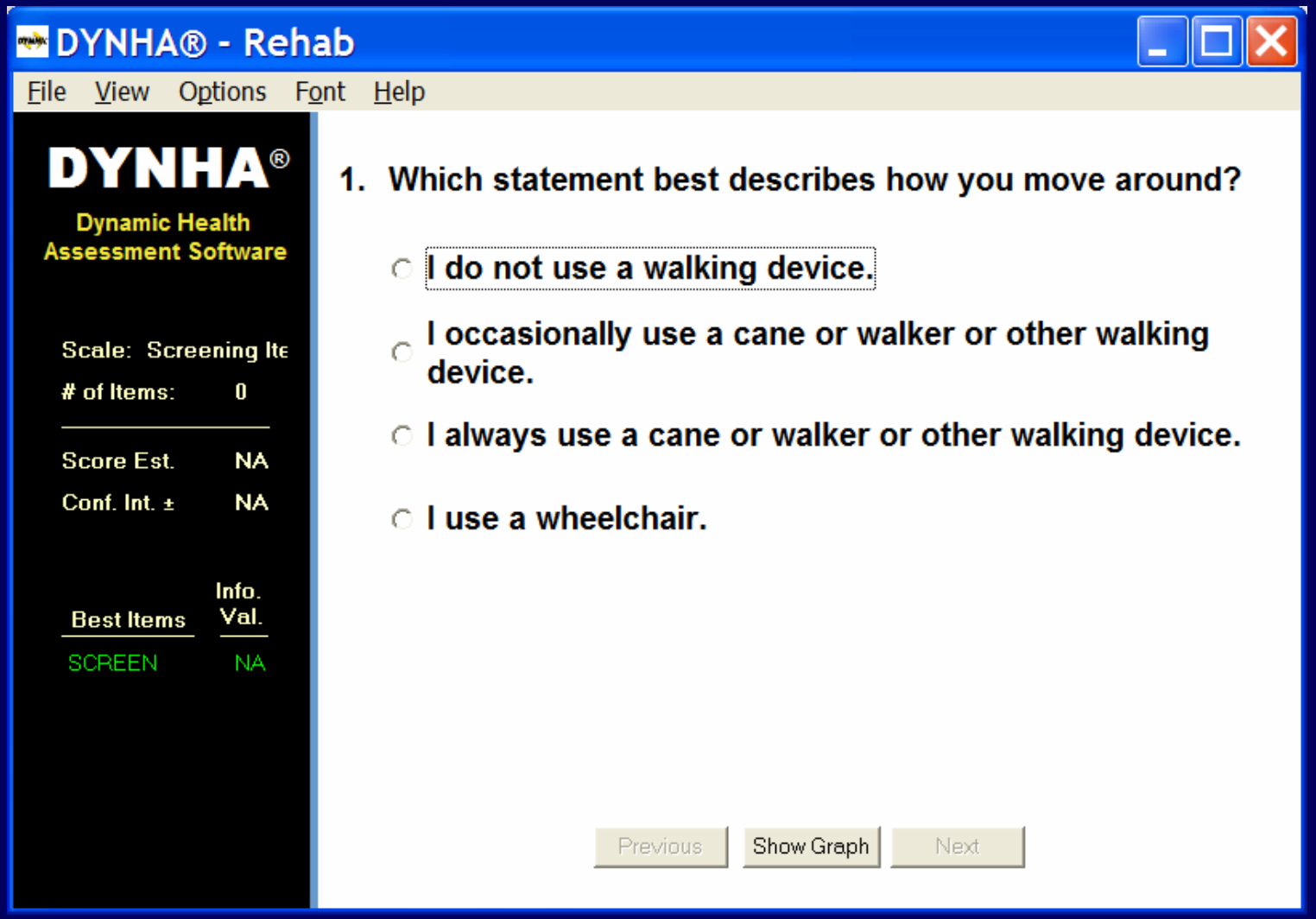
Computerized Dynamic Physical Activity Assessment

- Demo uses 101 physical activity items
- Preliminary calibrations from BU Project
- Dynamic Health Assessment (DYNHA®) Software
- Evaluation using “real data” simulation method

Movement & Physical Activity Item Set (N=101)

- 5 FIM items
- 9 SF-36 physical functioning items
- 11 MDS-PAC items
- 8 MDS items
- 5 OASIS items
- 63 new items (16 device, 8 wheelchair)

Rehabilitation Case Study # 1

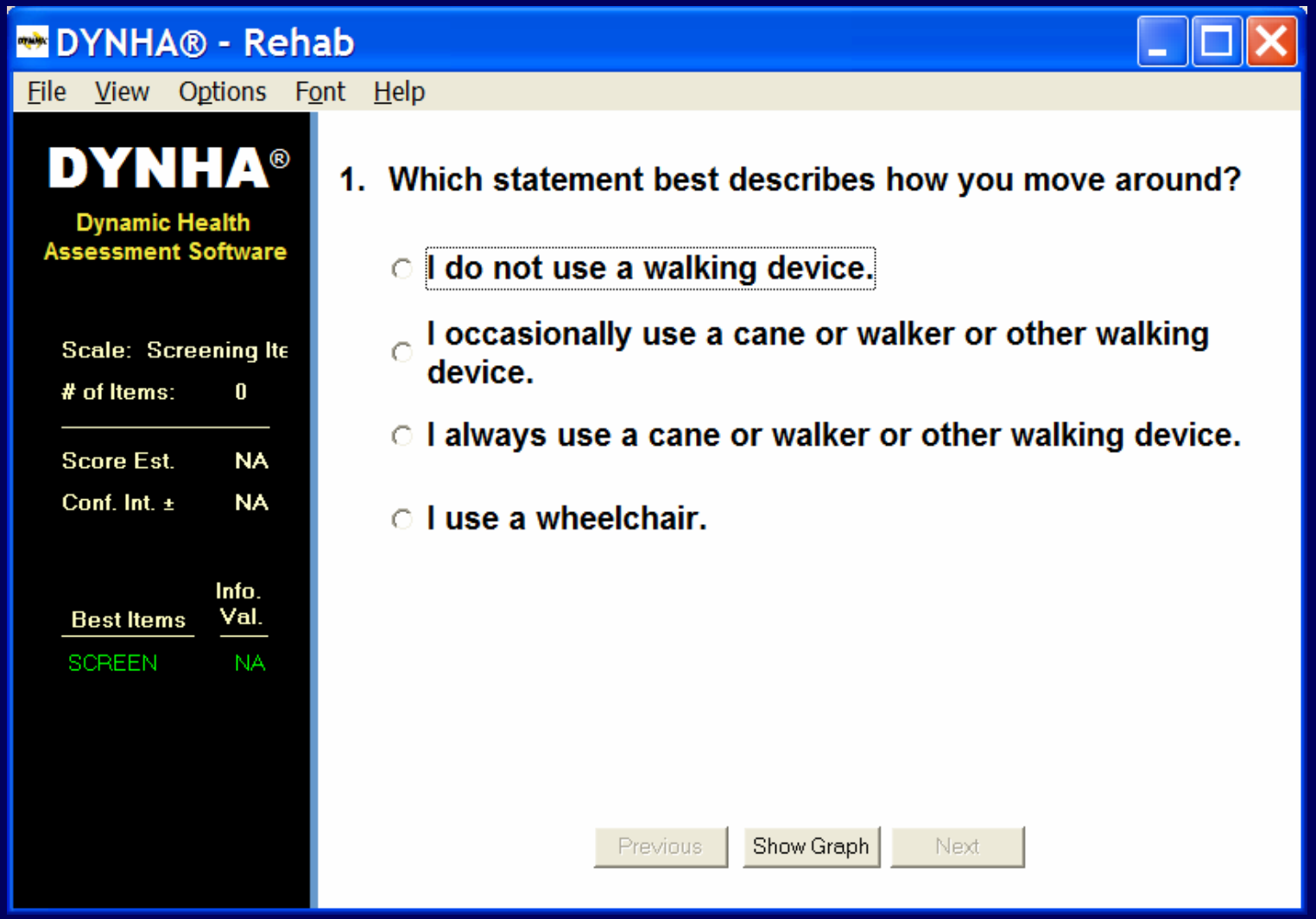

DYNHA® - Rehab
File View Options Font Help
DYNHA®
Dynamic Health Assessment Software
Scale: Screening Ite
# of Items: 0
Score Est. NA
Conf. Int. ± NA
Best Items
Info. Val.
SCREEN NA
1. Which statement best describes how you move around?
I do not use a walking device.
I occasionally use a cane or walker or other walking device.
I always use a cane or walker or other walking device.
I use a wheelchair.
Previous
Show Graph
Next

How Well Do the Results Agree?

	CAT DYNHA	Total Item Pool
Physical Activity Score	37.5	35.5
95% confidence interval	2.9	1.1
Respondent burden (items)	5	60

Source: Haley SM, Coster WJ, Andres PL, Kosinski M, Ni P. *Arch Phys Med Rehabil* 2004, Apr; 85(4): 661-6.

Rehabilitation Case Study # 2

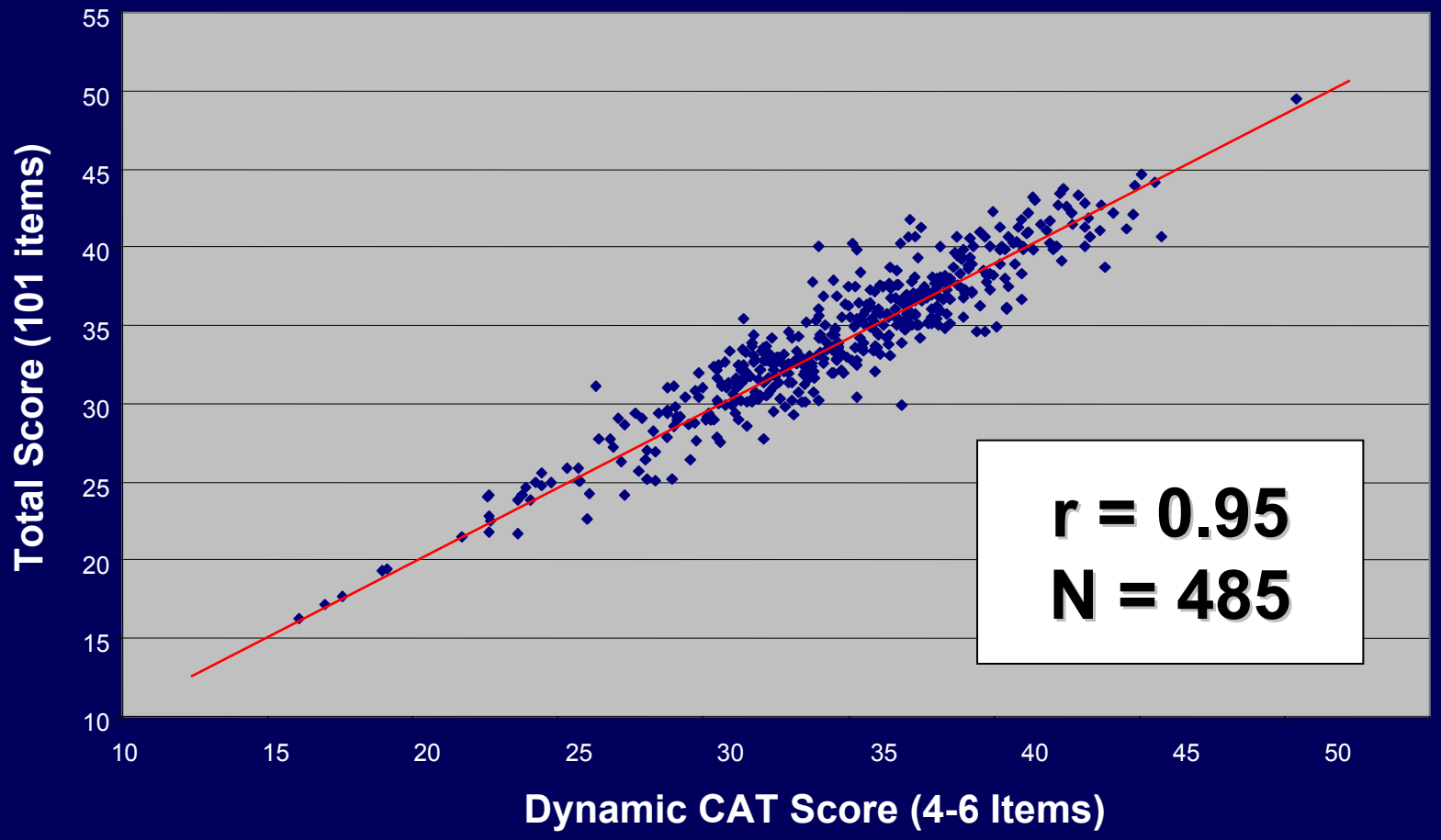

DYNHA® - Rehab
File View Options Font Help
DYNHA®
Dynamic Health Assessment Software
Scale: Screening Ite
# of Items: 0
Score Est. NA
Conf. Int. ± NA
Best Items
Info. Val.
SCREEN NA
1. Which statement best describes how you move around?
I do not use a walking device.
I occasionally use a cane or walker or other walking device.
I always use a cane or walker or other walking device.
I use a wheelchair.
Previous
Show Graph
Next

How Well Do the Results Agree?

	CAT DYNHA	Total Item Pool
Physical Activity Score	31.0	32.1
95% confidence interval	2.8	1.1
Respondent burden (items)	4	44

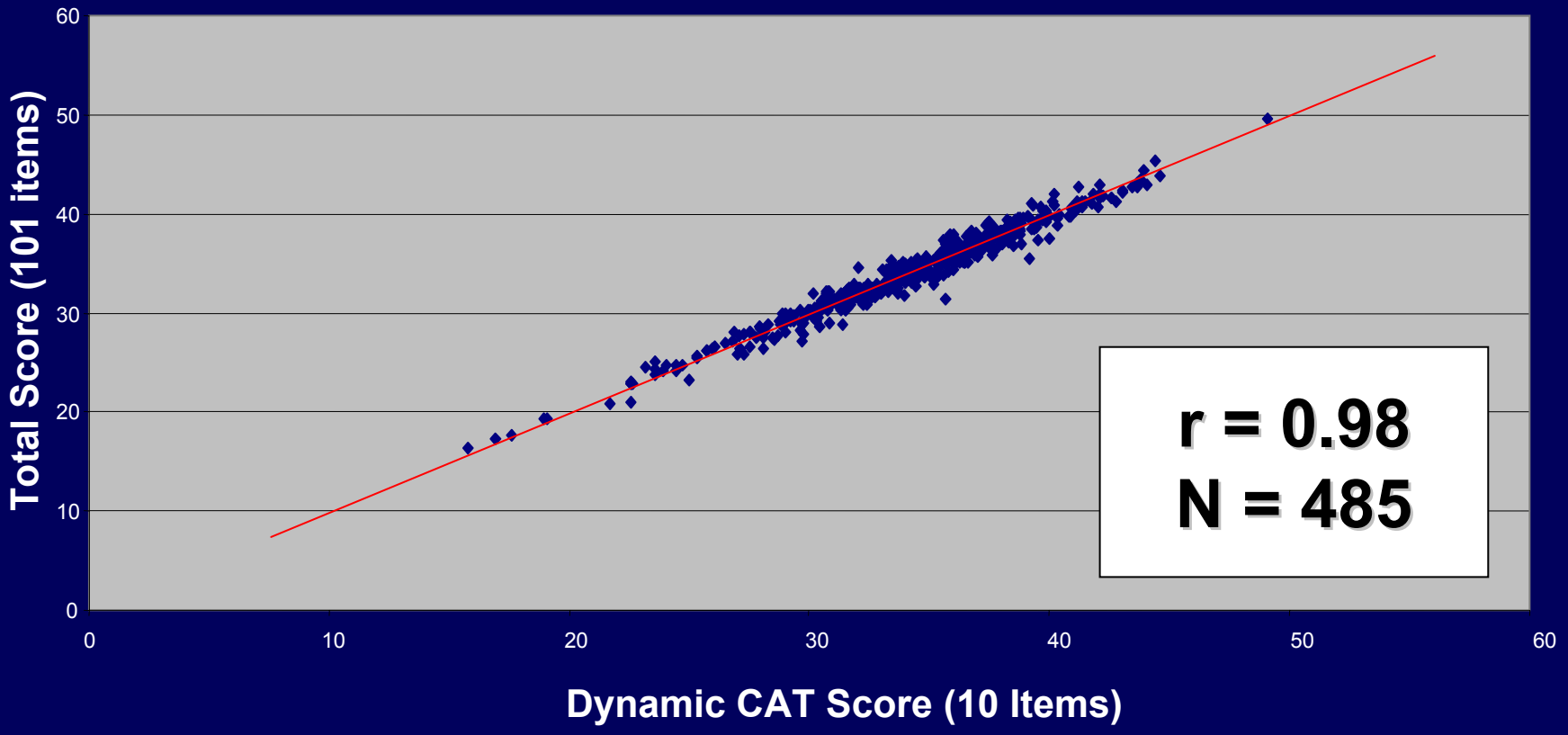
Source: Haley SM, Coster WJ, Andres PL, Kosinski M, Ni P. *Arch Phys Med Rehabil* 2004, Apr; 85(4): 661-6.

Plot of CAT versus Total Item Pool Estimates, AM-PAC Physical Activity

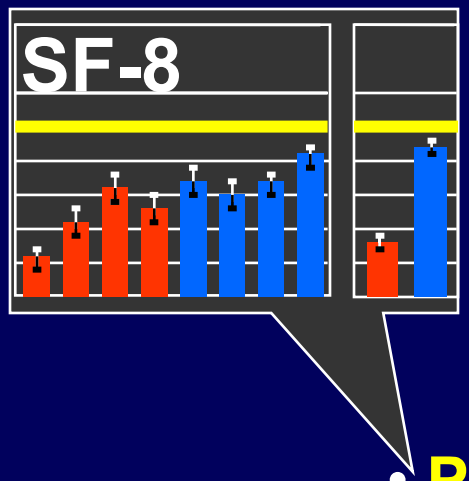


Haley SM, Coster WJ, Andres PL, Kosinski M, Ni P. *Arch Phys Med Rehabil* 2004, Apr; 85(4): 661-6.

Plot of CAT versus Total Item Pool Estimates, AM-PAC Physical Activity



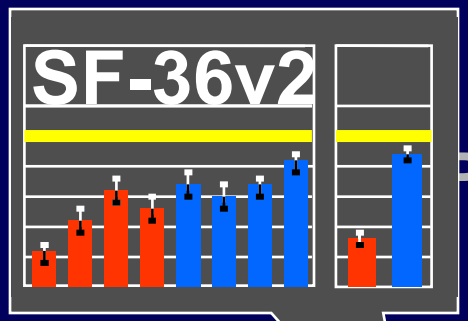
Haley SM, Coster WJ, Andres PL, Kosinski M, Ni P. *Arch Phys Med Rehabil* , 2004 Apr; 85(4): 661-6.



Standardizing Metrics and Matching Methods to Applications

- **Population level – What is normal health? (Shortest possible survey: one item per domain)**
- **Clinical trials & outcomes research – What work best?**
- **Patient-level screening & monitoring - How to improve decision-making?**

Standardizing Metrics and Matching Methods to Applications



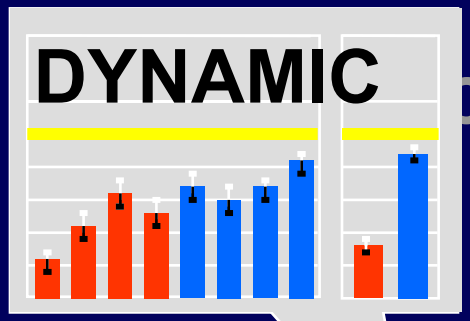
Population level – What is normal health?
(*Shortest possible survey: one item per domain)*

- **Clinical trials & outcomes research – What treatments work best?**
  (*Group-level standards of reliability and validity: multi-item scales)*

- **Patient-level screening & monitoring - How to improve decision-making?**

Standardizing Metrics and Matching Methods to Applications

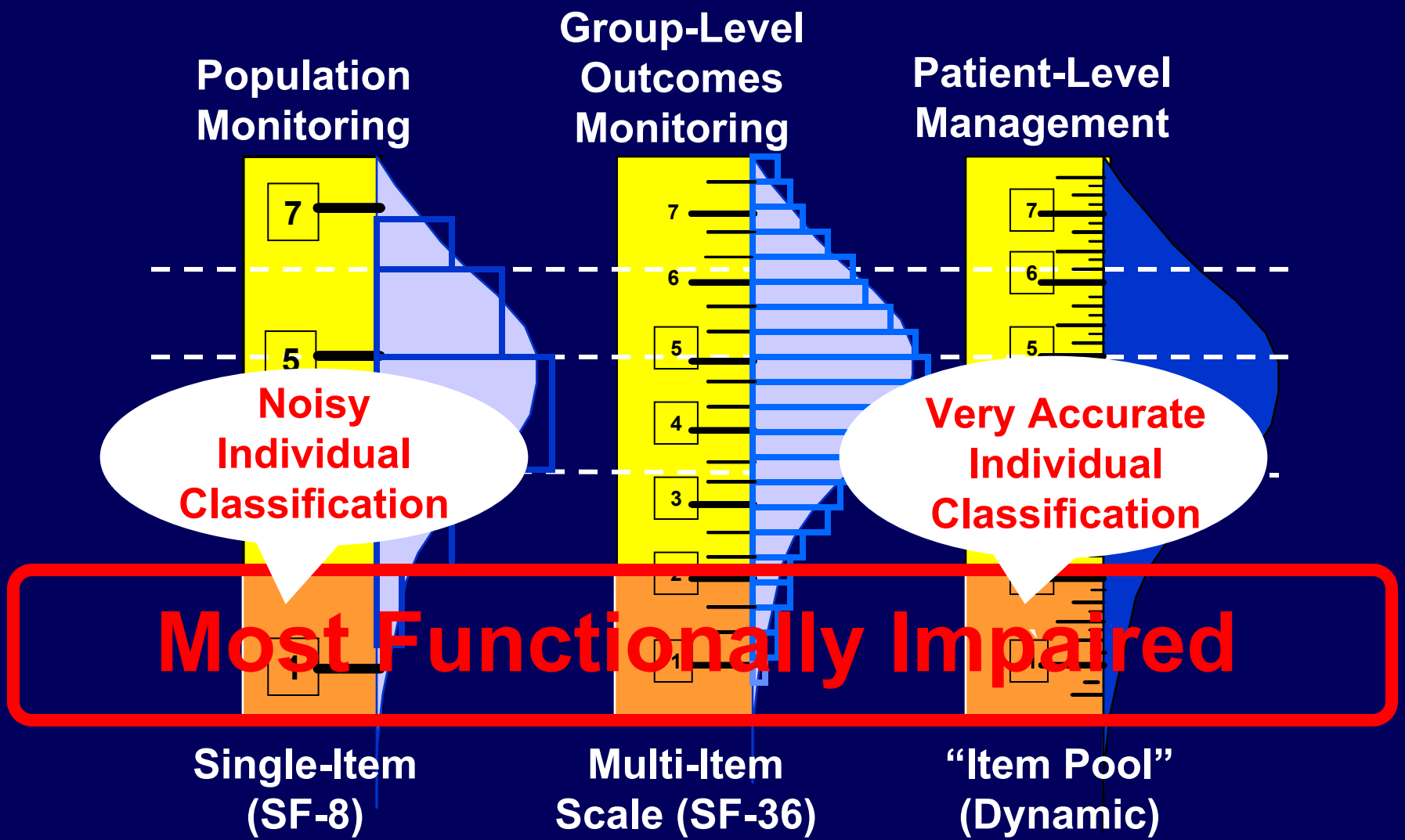
- Population level – What is normal health?
  (*Shortest possible survey)*



- Clinical trials & outcomes research –
  What work best?
  (*Group-level standards of reliability and validity: multi-item scales)*

- **Patient-level screening & monitoring -**
  How to improve decision-making?
  (Individual patient-*level standards of reliability and validity: CAT assessments)*

Matching Methods to Requirements of Each Application



What are the Advantages of Dynamic Assessments?

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

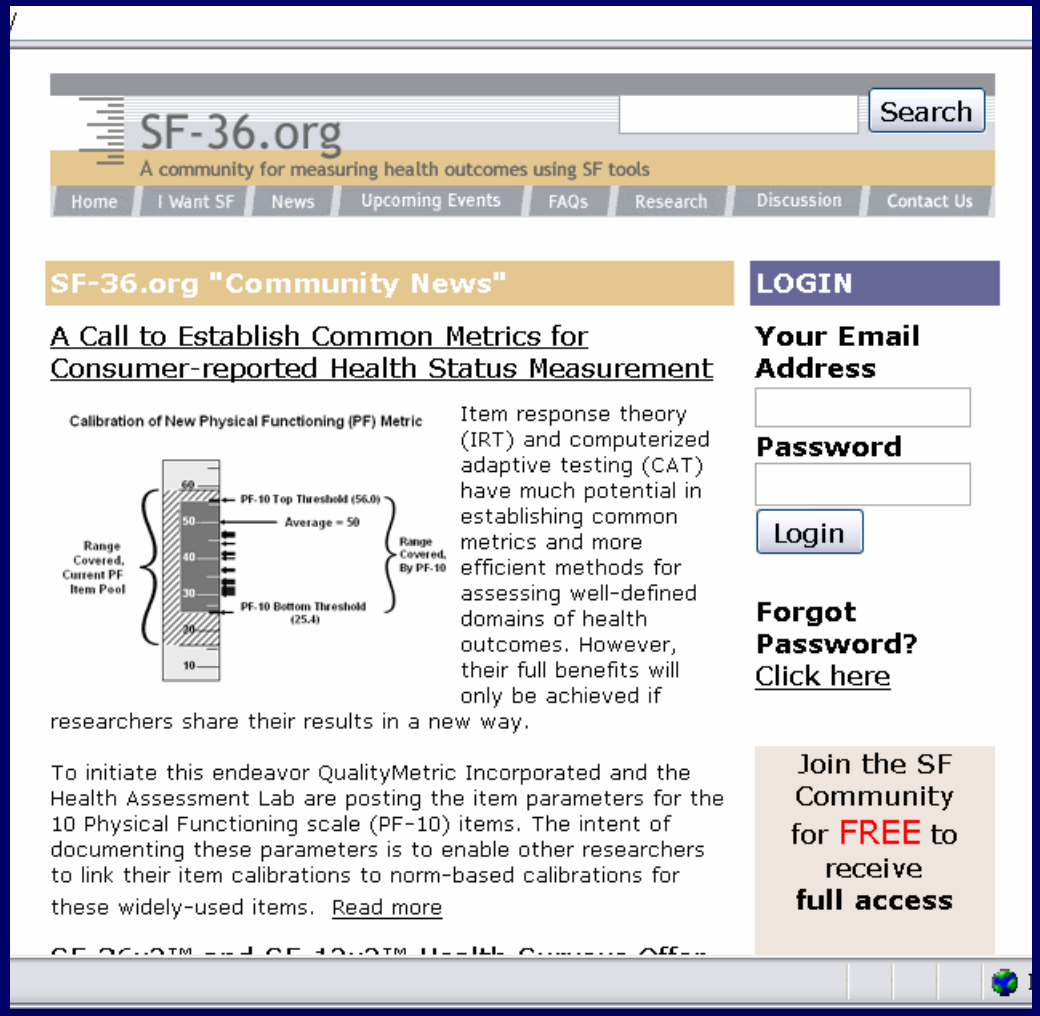
MEDICAL OUTCOMES TRUST

HAL Health Assessment Lab

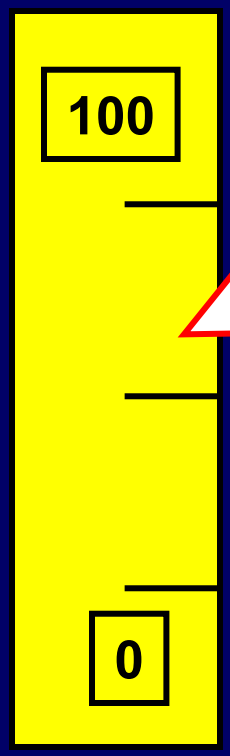
QM QUALITYMETRIC INCORPORATED

Public Documentation of Item Calibrations on the Internet

Physical Functioning (PF) Item Parameters Documented on Internet: www.sf-36.org



Physical Functioning (PF-10) Item Parameters Documented on the Internet: www.sf-36.org



PF-10

PF Item	Abbreviated Content	Threshold 1	Threshold 2
01	Vigorous activities	-0.612	0.602
02	Moderate activities	-1.626	-0.603
03	Lift/carry groceries	-1.899	-0.895
04	Climb several flights of stairs	-1.288	-0.246
05	Climb one flight of stairs	-1.927	-0.995
06	Bend/kneel/stooping	-1.598	-0.312
07	Walking more than a mile	-1.703	-0.381
08	Walking several blocks	-1.409	-0.857
09	Walking one block	-1.980	-1.326
10	Bathing or dressing	-2.458	-1.865

Note: Slopes = 2.558; 1998 representative US sample (N = 6,303)
For more information go to www.sf-36.org

Collecting and Processing HR-QOL Data

Personal Interview

Telephone Interview

Self-Administered Questionnaire

Computerized Personal Telephone Interview

Internet

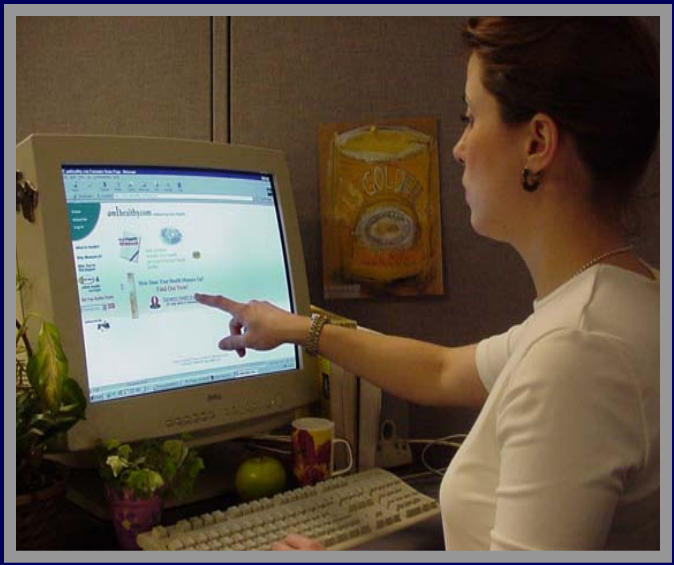
Handhelds

Articles Documenting the Development and Testing of the Headache Impact Test (HIT), *Quality of Life Research*, 2003, 12, 887-1012

1. Bjorner et al., The Feasibility of Applying Item Response Theory to Measures of Migraine Impact: A Re-analysis of Three Clinical Studies
2. Kosinski et al., The Responsiveness of Headache Impact Scales Scored Using 'Classical' and 'Modern' Psychometric Methods: A Re-Analysis of Three Clinical Trials
3. Bjorner et al., Calibration of an Item Pool for Assessing the Burden of Headaches: An Application of Item Response Theory to the Headache Impact Test (HIT™)
4. Ware et al., Applications of Computerized Adaptive Testing (CAT) to the Assessment of Headache Impact
5. Bayliss et al., A Study of the Feasibility of Internet Administration of a Computerized Health Survey: The Headache Impact Test (HIT™)
6. Kosinski et al., A 6-item Short-Form Survey for Measuring Headache Impact: The HIT-6™
7. Gandek et al., Translating the Short-Form Headache Impact Test (HIT-6) in 27 Countries: Methodological and Conceptual Issues
8. Bjorner et al., Using Item Response Theory to Calibrate the Headache Impact Test (HIT™) to the Metric of Traditional Headache Scales
9. Turner-Bowker et al., Usefulness of the SF-8™ Health Survey for Comparing the Impact of Migraine and Other Conditions

Acknowledgements: NIH and DOE Grant Support

Title	Target Population	Agency	Institute
Computerized Adaptive Version of the PEDI (Phase I)	Pediatric	National Institutes of Health	National Institute of Child Health & Human Development (NICHD)
Functional Health CAT in Chronic Kidney Disease (Phase I)	Chronic Kidney Disease	National Institutes of Health	National Institute of Diabetes & Digestive & Kidney Disease (NIDDK)
Rehabilitation Research and Training Center on Measuring Rehabilitation Outcomes	Rehab (Adult)	Department of Education	National Institute on Disability and Rehabilitation Research (NIDRR)
Functional Health CAT in Diabetes (Phase I)	Diabetes	National Institutes of Health	National Institute of Diabetes & Digestive & Kidney Disease (NIDDK)
Computerized Adaptive Assessment of Headache Impact (Phase I)	Headache	National Institutes of Health	National Institute of Neurological Disorders and Stroke (NINDS)
Dynamic Assessment of Pediatric Health and Functioning	Pediatric	National Institutes of Health	National Institute of Child Health & Human Development (NICHD)
Computerized Adaptive Version of the PEDI (Phase II)	Pediatric	National Institutes of Health	National Institute of Child Health & Human Development (NICHD)
Computerized Adaptive Assessment of Asthma Impact (Phase I)	Asthma	National Institutes of Health	National Heart, Lung, and Blood Institute



Computerized Dynamic Health Assessment

Free demos:

www.amlhealthy.com
www.headachetest.com
www.qualitymetric.com