

Choosing Patient-Reported Outcomes and Measurement Methods for Team Based Health Care

John E. Ware, Jr., PhD

Professor, Department of Quantitative Health Sciences, UMass Medical School
Chief Science Officer, John Ware Research Group, Watertown, MA 02472

Integrated Health Care Conference
Arizona State University Doctor of Behavioral Health Program
Scottsdale, AZ - March 5-7, 2018

Quality and Outcomes

“The best measure of quality is not how well or how frequently a medical service is given, but how closely the result approaches the fundamental objectives of prolonging life, relieving distress, restoring function, and preventing disability.”

Lembcke, 1952

Lembcke , PA. Measuring the quality of medical care through vital statistics based on hospital service areas: 1. Comparative study of appendectomy rates. *Am J Public Health* 1952; 42: 276-86.

**Fee-for-outcome,
not fee-for-service,**

Background

- Proven value of patient-reported outcome (PRO) measures (even short-forms) in prediction and monitoring health care
- Increased understanding of functional health and well-being has led to much better PRO tools
- Standardization (concepts represented, range measured, underlying metrics) is progressing across applications

Approach

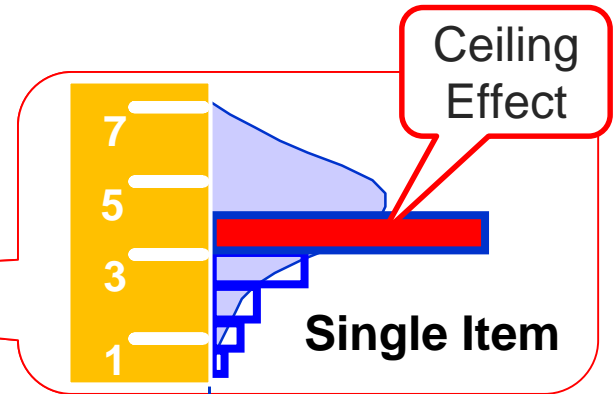
- More actionable integration of disease-specific and generic measures
- More practical Internet-based data capture
- Better-matching of measurement methods and applications
- Directly comparable metrics across methods
- Better single-item measures

What Are Some Distinct Applications?

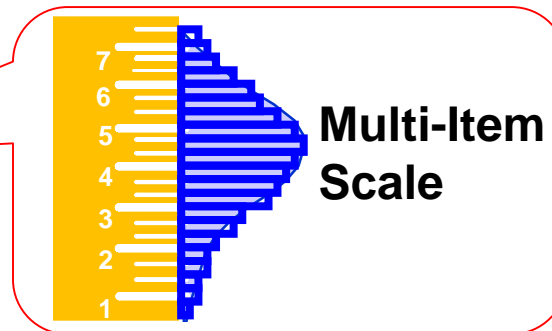
- Population monitoring
- Group-level outcomes monitoring
- Patient-level screening and monitoring

Different Outcome Monitoring Applications and Matching Methods

- Population



- Group-level



- Individual patient-level



Quality of Life (QoL)

- Community
- Education
- Family Life
- Friendships
- **Health**
- Housing
- Marriage
- Nation
- Neighborhood
- Self
- Standard of Living
- Work

Source: Campbell, A. *The sense of well-being in America: Recent patterns and trends*. New York: McGraw-Hill. 1981

World Health Organization Definition of Health

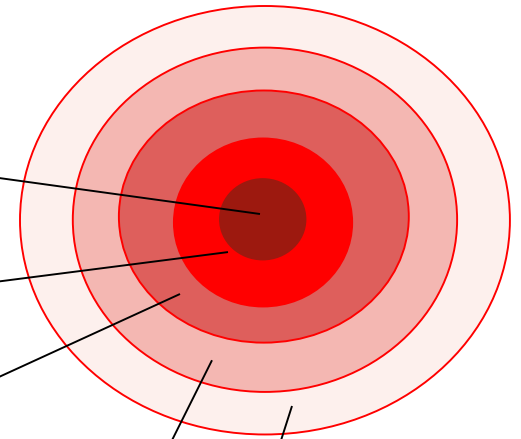
“**Health** is a state of complete **physical**, **mental**, and **social** well-being and not merely the absence of disease or infirmity”

WHO, 1948

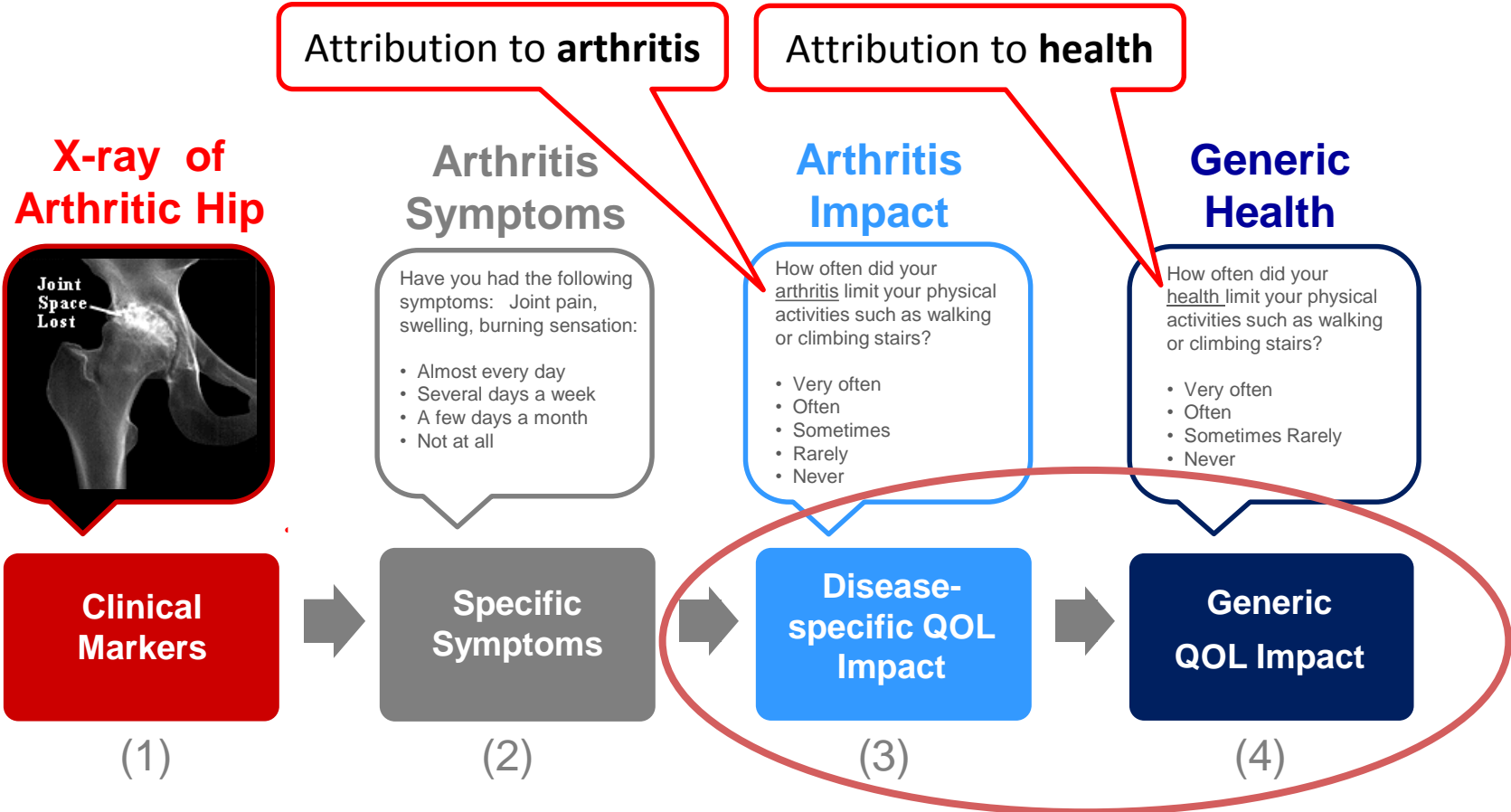
Health Can Be Measured Using Different Operational Definitions:

- Bodily structure & function
- Specific symptoms
- What you do/are able to do – functioning
- How you feel – subjective ill- and well-being (+ and -)
- What you say it is – personal evaluation

QOL



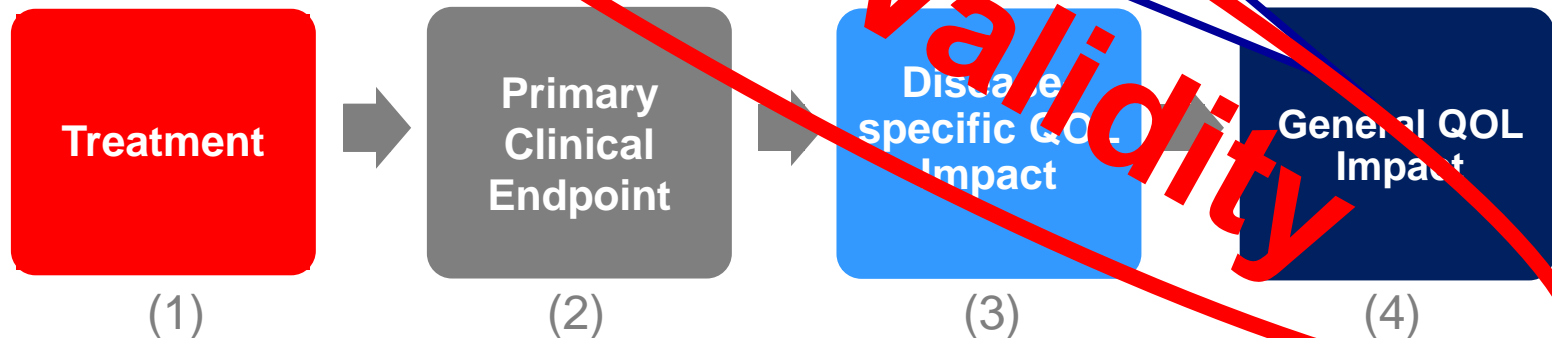
Continuum of Disease-specific and General Health Measures



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

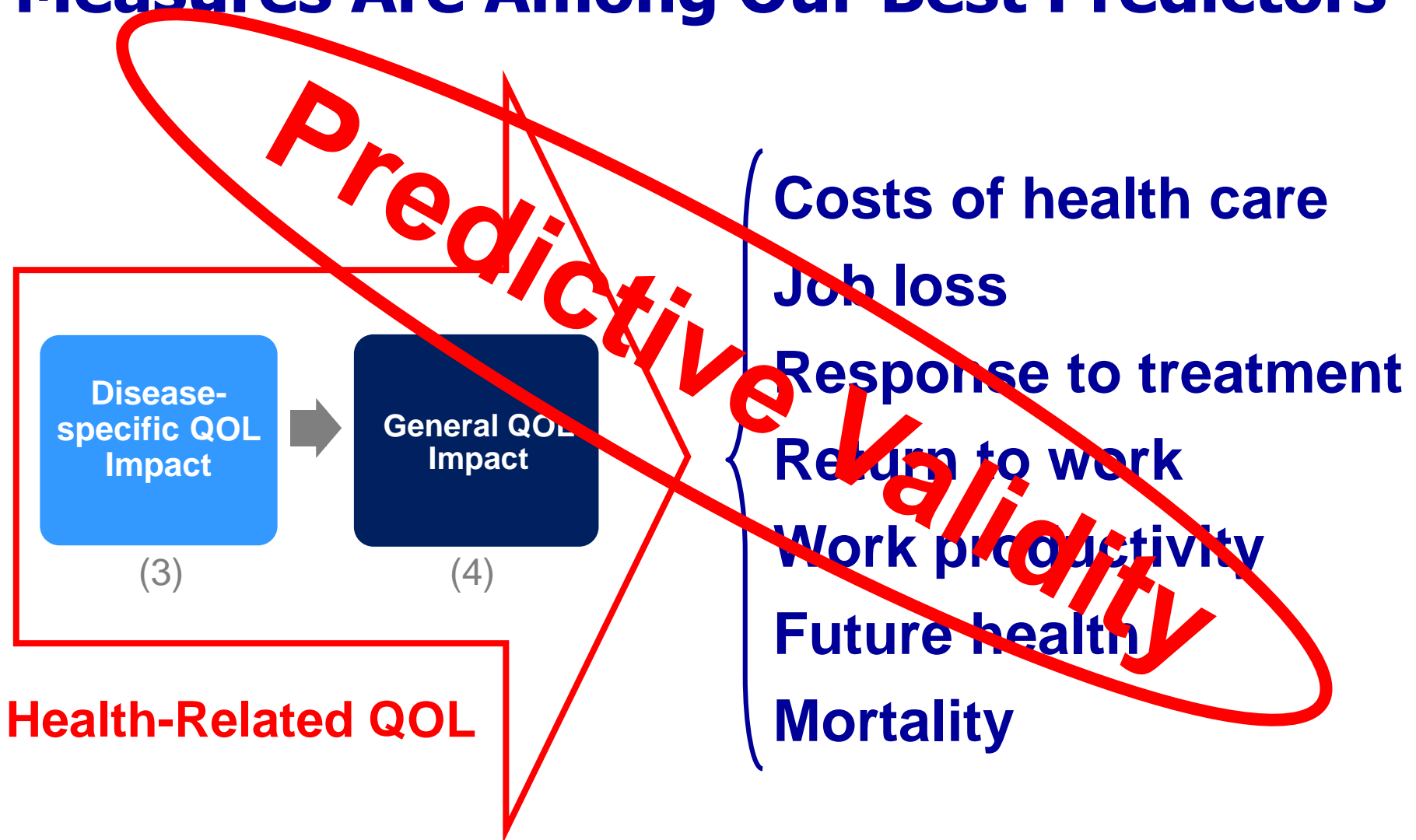
Generic QOL Measures Respond to 87% of Clinically Efficacious Treatments

In the first 17 years of well-controlled pharmaceutical RCTs, across 14 therapeutic areas, SF-36 summary measures responded significantly when primary clinical endpoints responded 87% of the time; 58% > MID

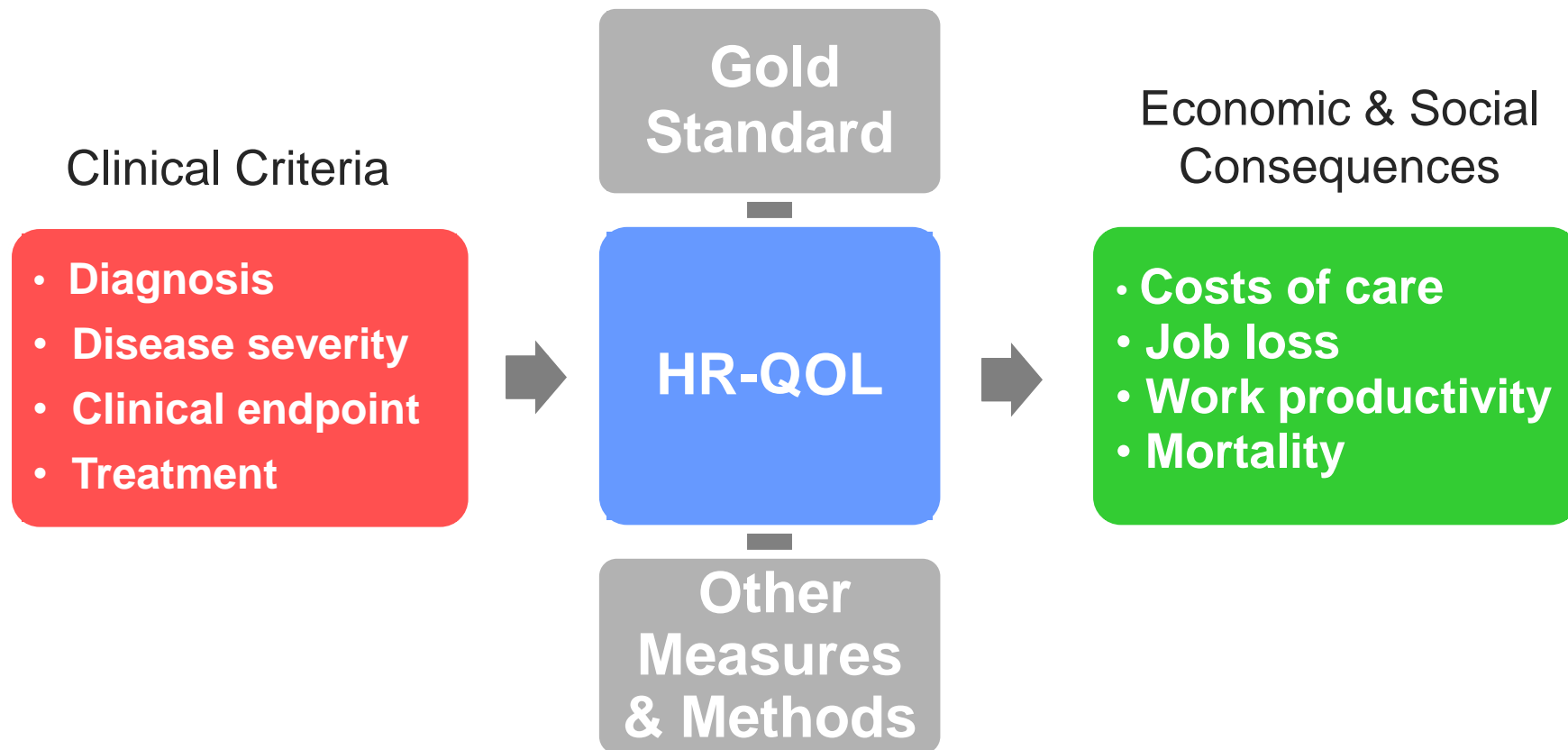


Reference: Frenkl D and Ware JE. "Patient-reported functional health and well-being outcomes with drug therapy: A systematic review of randomized trials using the SF-36 Health Survey". *Medical Care*, 2014, 52(5):439-445.

Generic and Disease-specific QOL Measures Are Among Our Best Predictors



Measurement Validation & Interpretation



Adapted from: Ware JE, Jr. and Keller SD: Interpreting general health measures, in: *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia, PA: Lippincott-Raven Publishers; 1995: Chapter 47.

40-year Trends in Health Domains Represented in Widely-used Surveys

Concepts/Domains	Psychometric										Utility			
	SIP	HIE	NHP	COOP	DUKE	MOS FWBP	MOS SF-36	EORTC QLQ-C30	PROMIS 29-57		QWB	EQ-5D	HUI	SF-6D
Physical functioning (-)	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Social functioning (-)	•	•	•	•	•	•	•	•	•	•	•			•
Role functioning (-)	•	•	•	•	•	•	•	•	•	•	•			•
Psychological distress (-)	•	•	•	•	•	•	•	•	•	•	•		•	•
Health perception (general)		•	•	•	•	•	•	•	•	•				
Pain (-)		•	•	•	•	•	•	•	•	•		•	•	•
Fatigue (-)		•	•		•	•	•	•	•	•	•			
Energy		•				•	•		•					•
Psychological well-being		•			•	•	•		•					•
Sleep	•		•		•	•		•			•			
Cognitive functioning (-)	•				•	•		•					•	
Quality of life		•		•		•		•			•			
Evaluated health outcome				•		•		•	•					

Psychometric

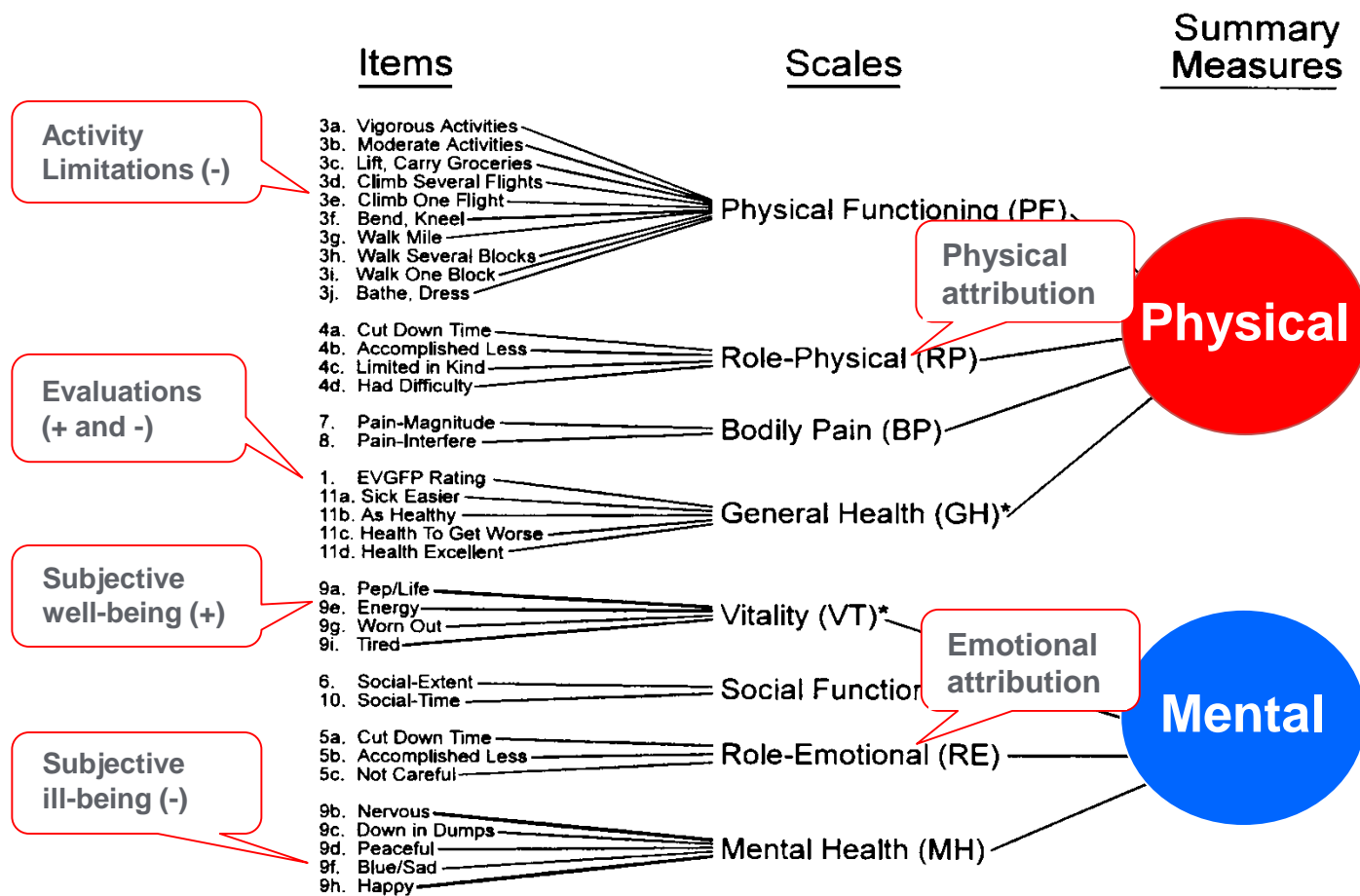
SIP =Sickness Impact Profile (1976) HIE=Health Insurance Experiment (1979) NHP=Nottingham Health Profile (1980) COOP=Dartmouth Function Charts (1987) DUKE=Duke Health Profile (1990) MOS FWBP=MOS Functioning & Well-Being Profile (1992) MOS SF-36=36-Item Short-Form Health Survey (1992) EORTC QLQ-30=European Organization for Research & Treatment of Cancer (1993) PROMIS=Patient Reported Outcomes Measurement Information System (2004-on)

Utility

QWB=Quality of Well-Being Scale (1973) EQ-5D= European Quality of Life Index (1990) HUI=Health Utility Index (1996) SF-6D=SF-36 Utility Index (2002)

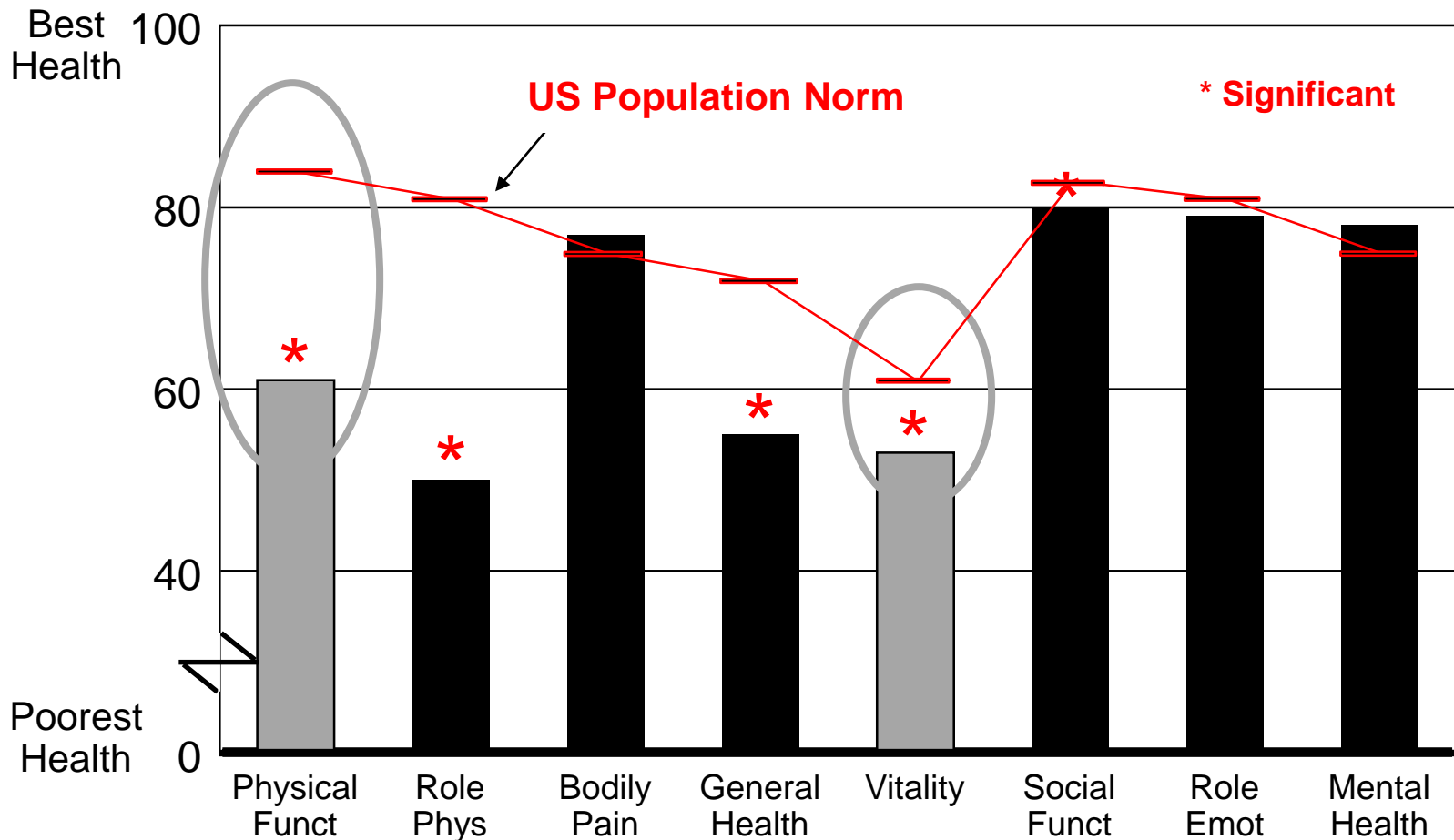
Source: Adapted from Ware 1987; Ware, 1995

MOS* Measurement Model, Operational Definitions and SF-36 Survey Items



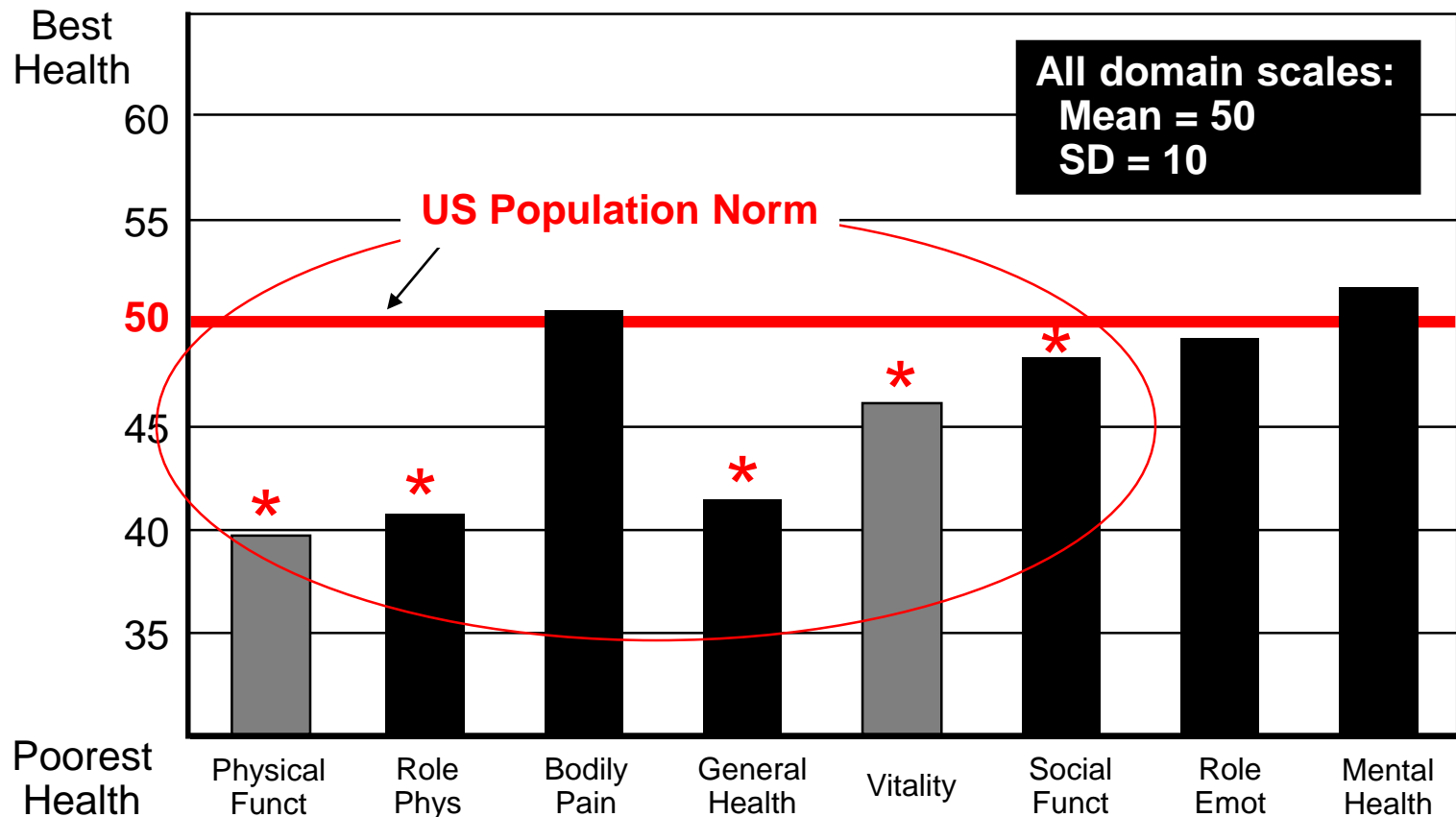
*Medical Outcomes Study (MOS) surveys spawned the SF-36, SF-12, SF-8 health surveys

Worst to Best (0-100) Scoring Simplified Health Profile Interpretation: Asthma RCT



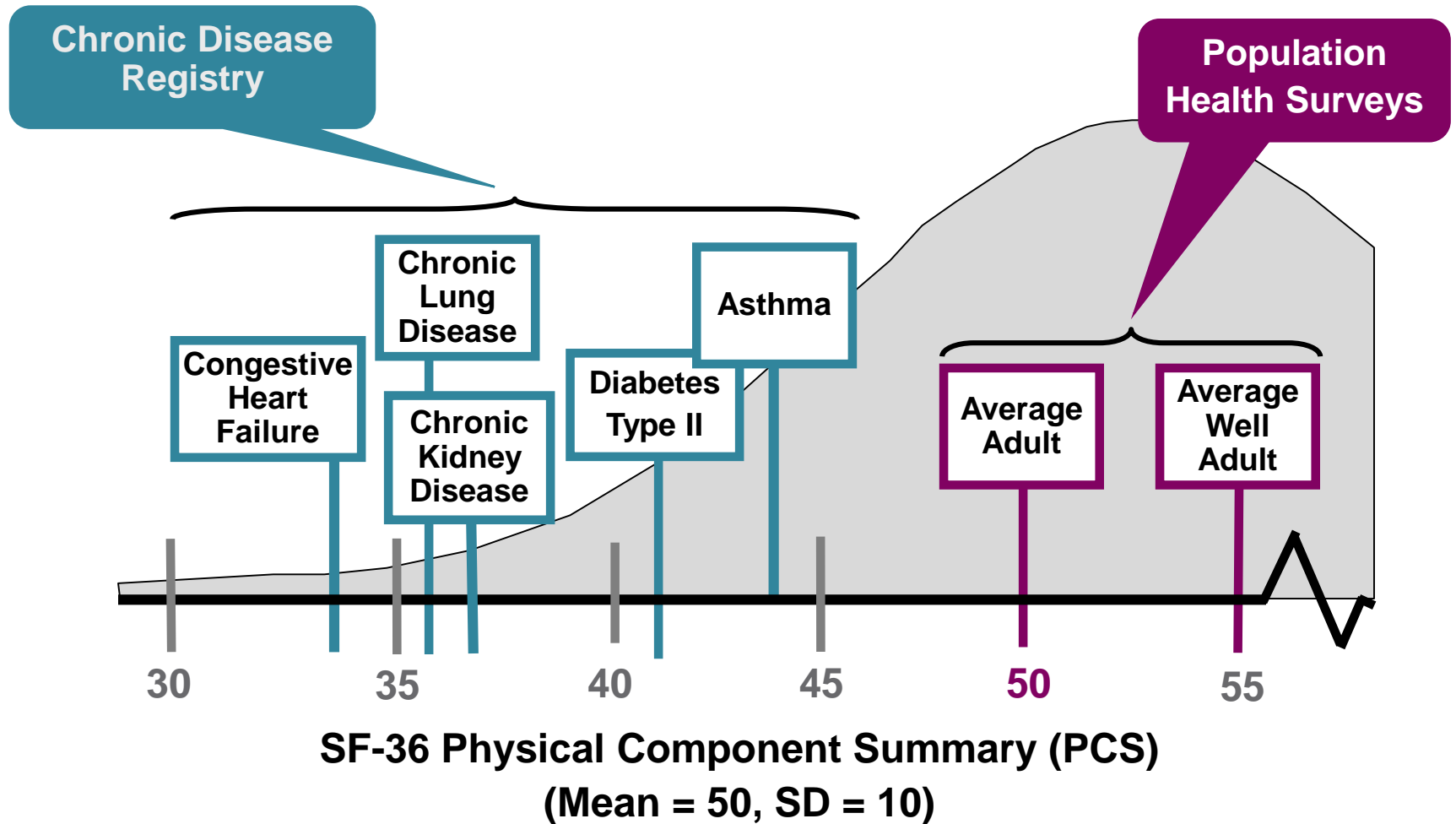
Source: Okamoto et al., Medical Outcomes Study Short-Form 36 (SF-36), *Annals of Allergy, Asthma & Immunology*, 1996

Norm-based Scoring of Profiles Makes Interpretation Easier: Asthma RCT

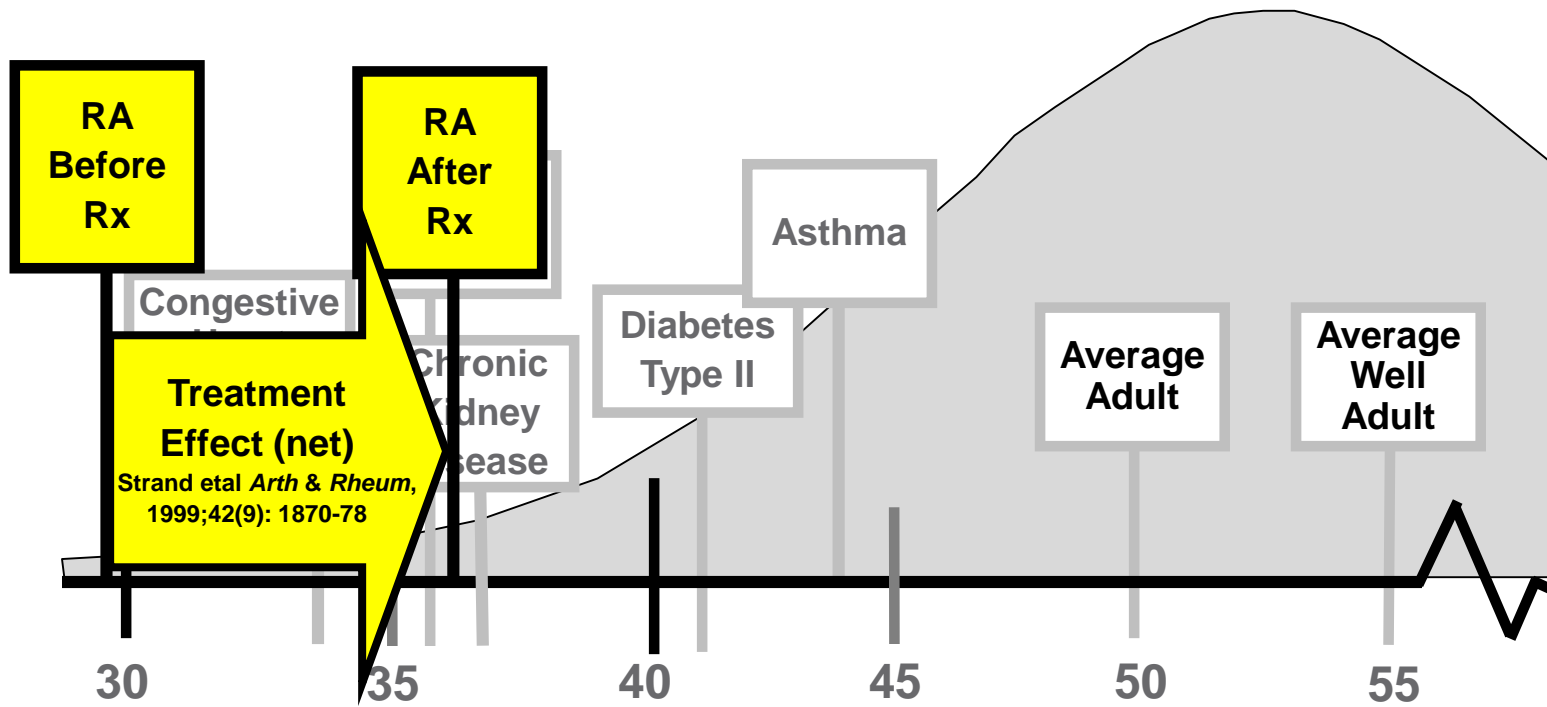


Source: Norm-based scoring of Okamoto et al., MOS Short-Form 36 (SF-36), *Annals of Allergy, Asthma & Immunology*, 1996

Standardized Summary Measures Link Disease Burden to Population Health



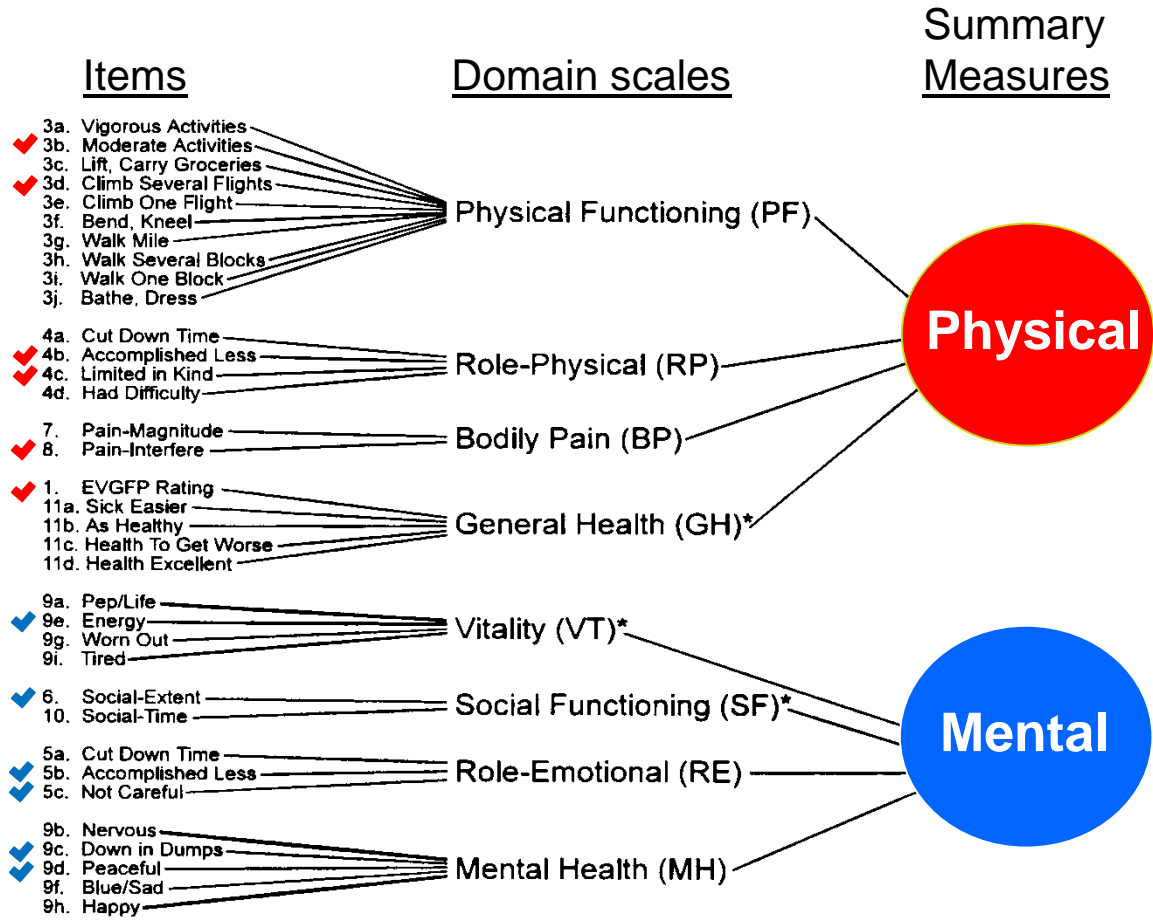
Advantage of Standardization: Linking Treatment Outcomes to Population Health



SF-36 Physical Component Summary (PCS)
(Mean = 50, SD = 10)

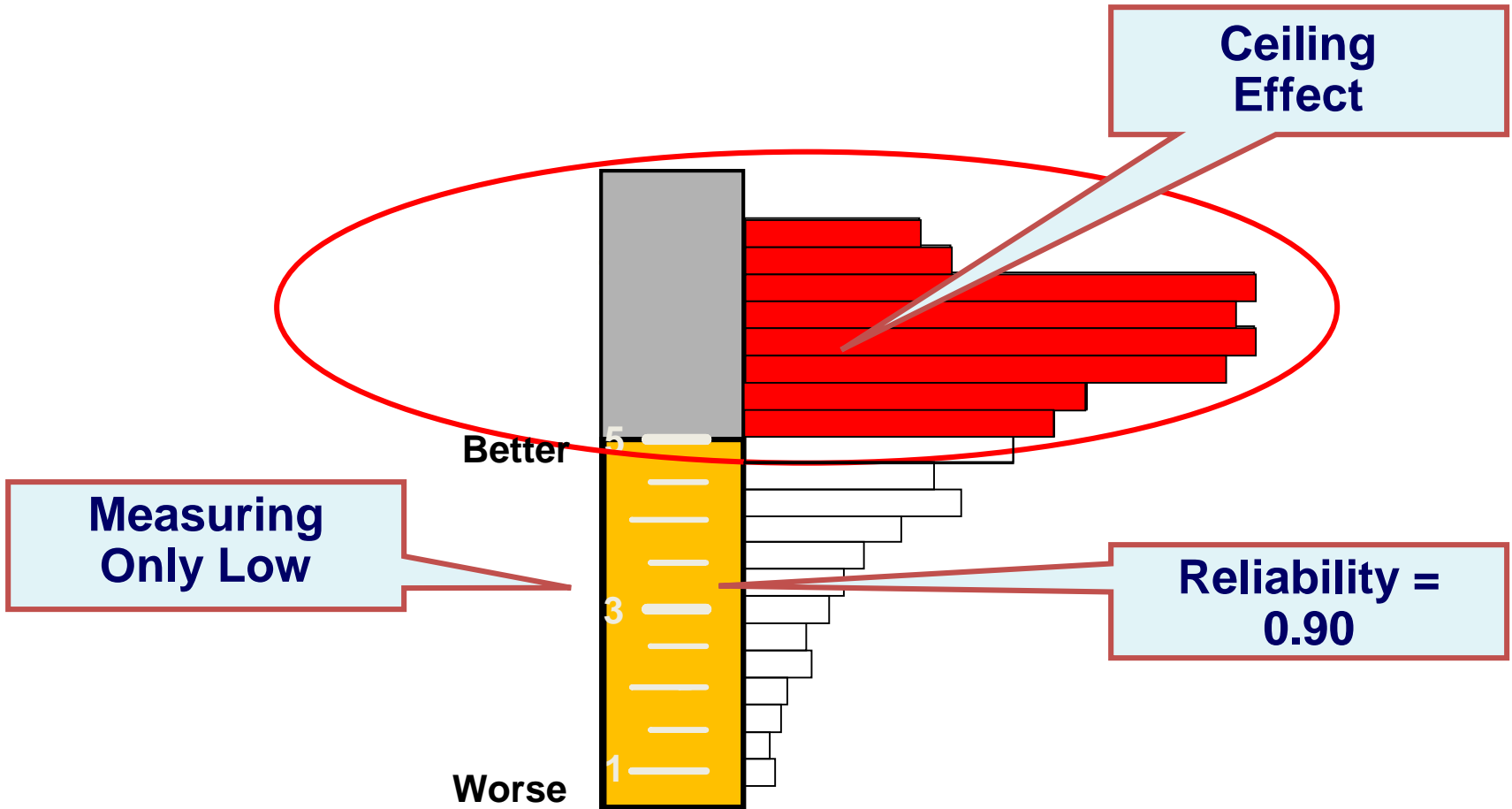
Strand et al *Arth & Rheum*,
1999;42(9): 1870-78

SF-12 Enabled Two-thirds Reduction in Respondent Burden

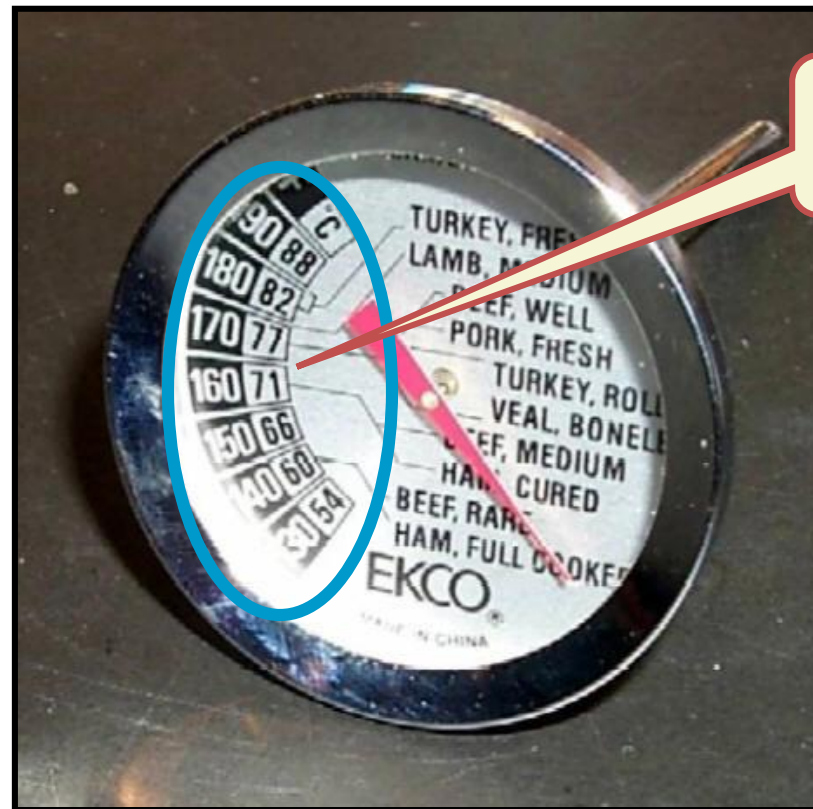


* Significant correlation with other summary measure.

Ceiling Effects

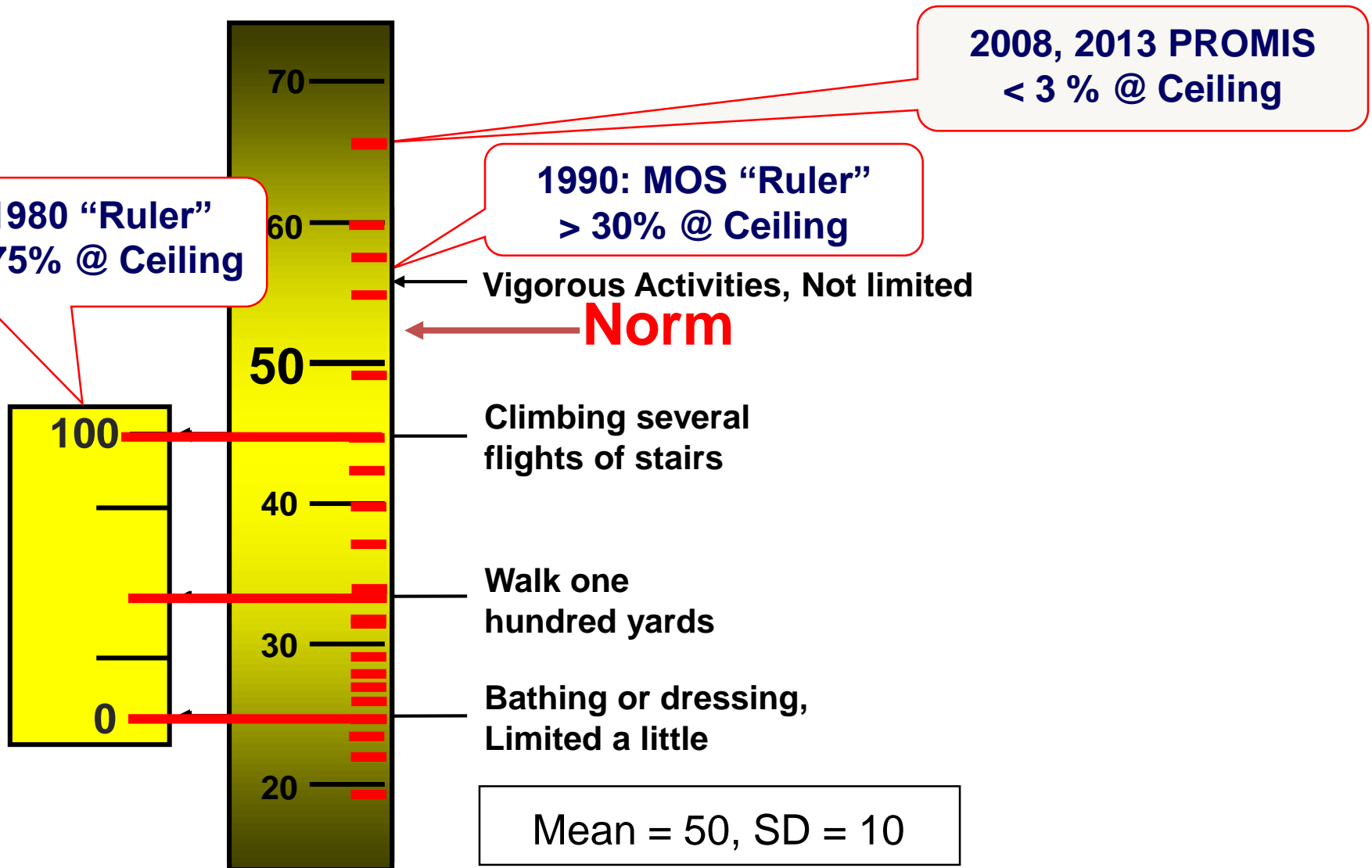


“Ceiling Effect” Example: Measuring Only High Temperatures



Cooking Thermometer

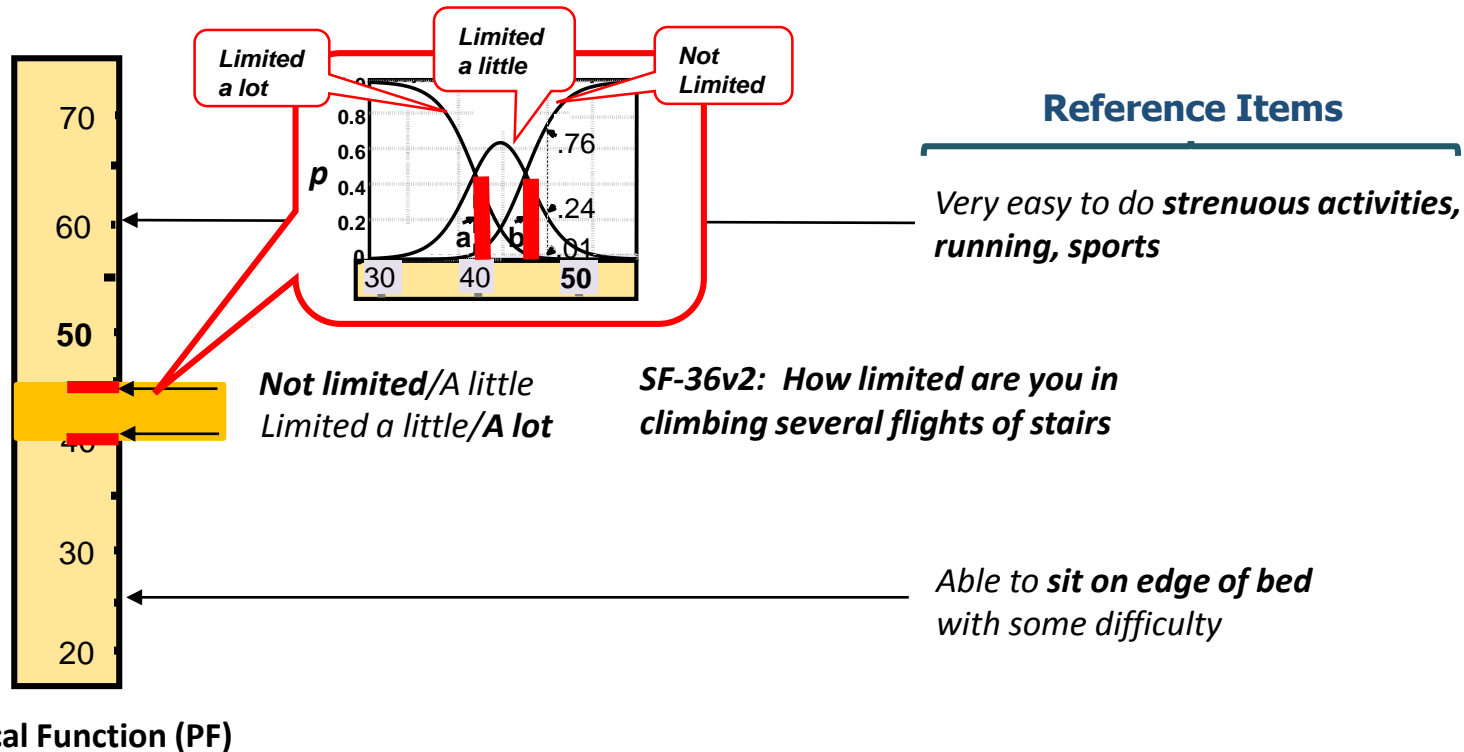
Brief History of Raising a Ceiling: Physical Function "Ruler"



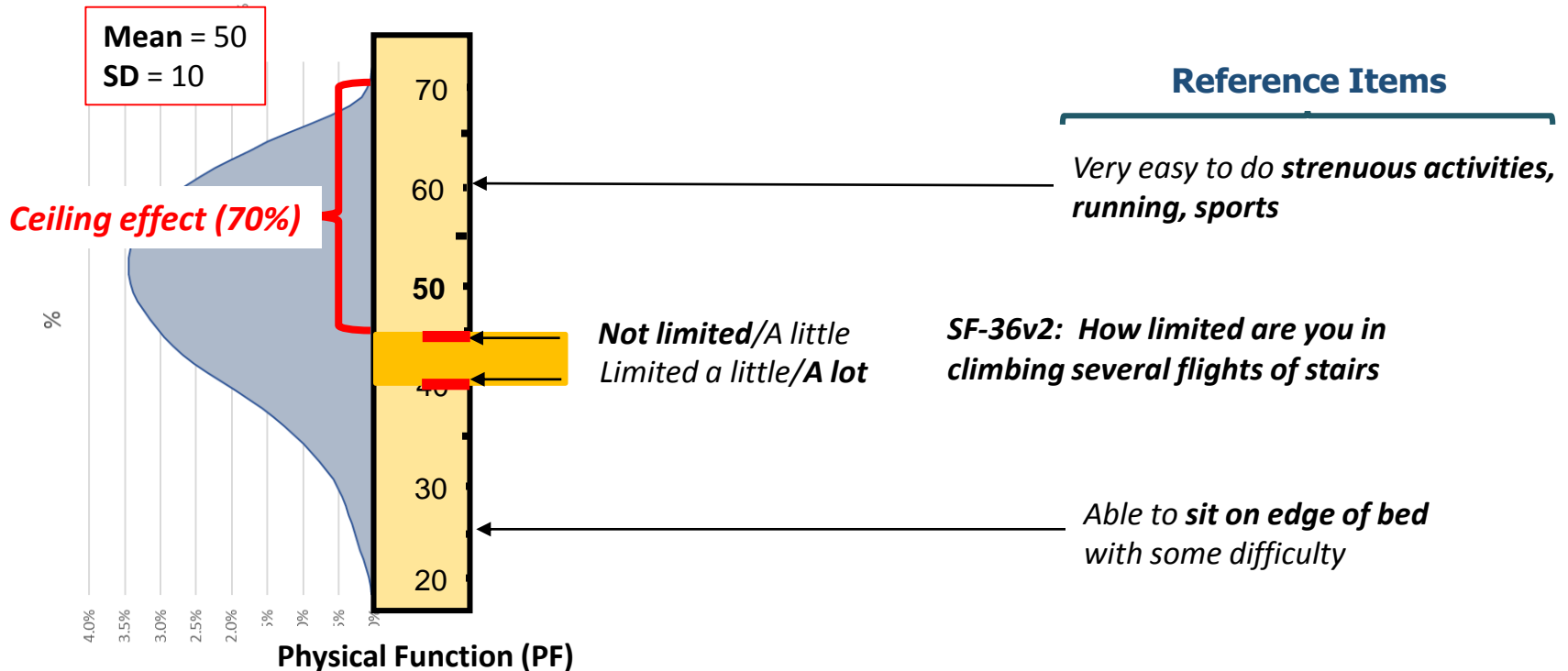
Item Response Theory (IRT) Models

Locate Marks on PF "Ruler"

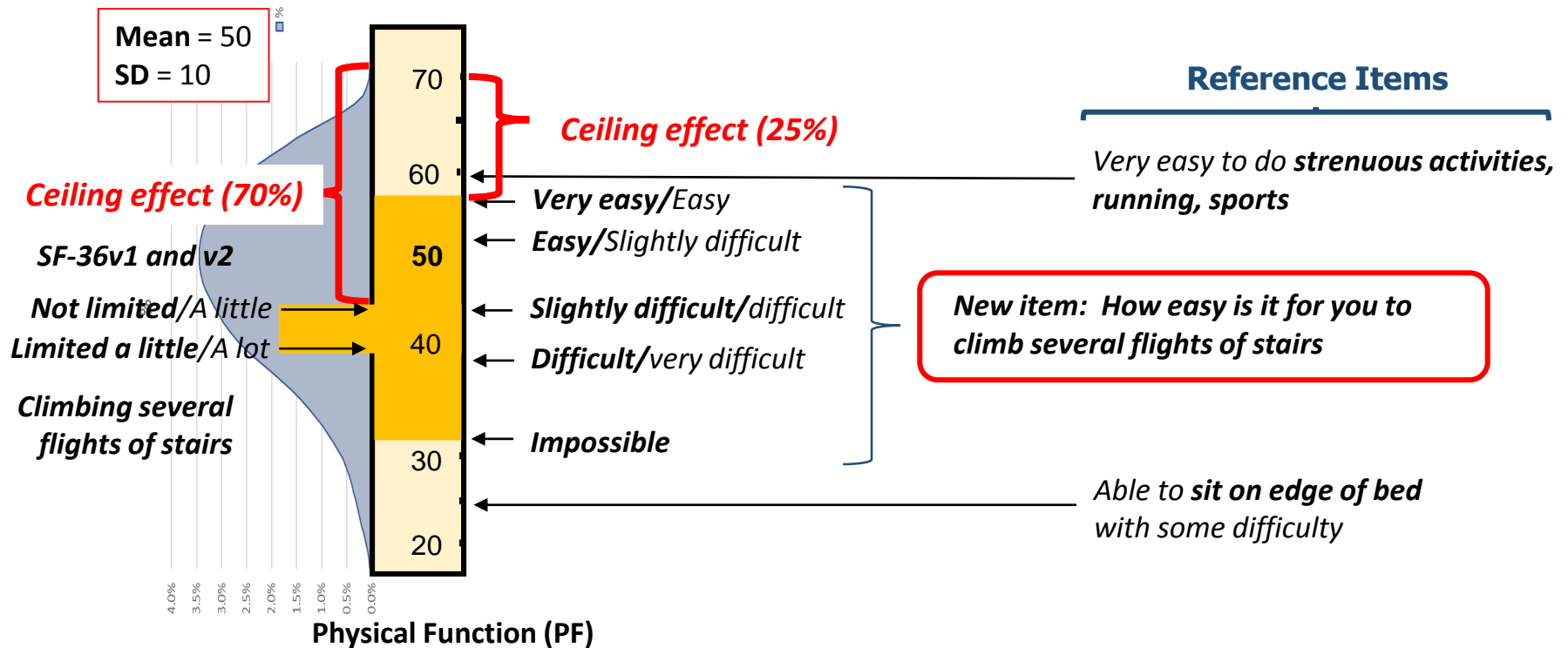
Mean = 50
SD = 10



Typical Single-item Response Categories Restrict Range of Physical Function Measurement (SF-36)



Extending Response Category Range Improves Physical Function Measurement

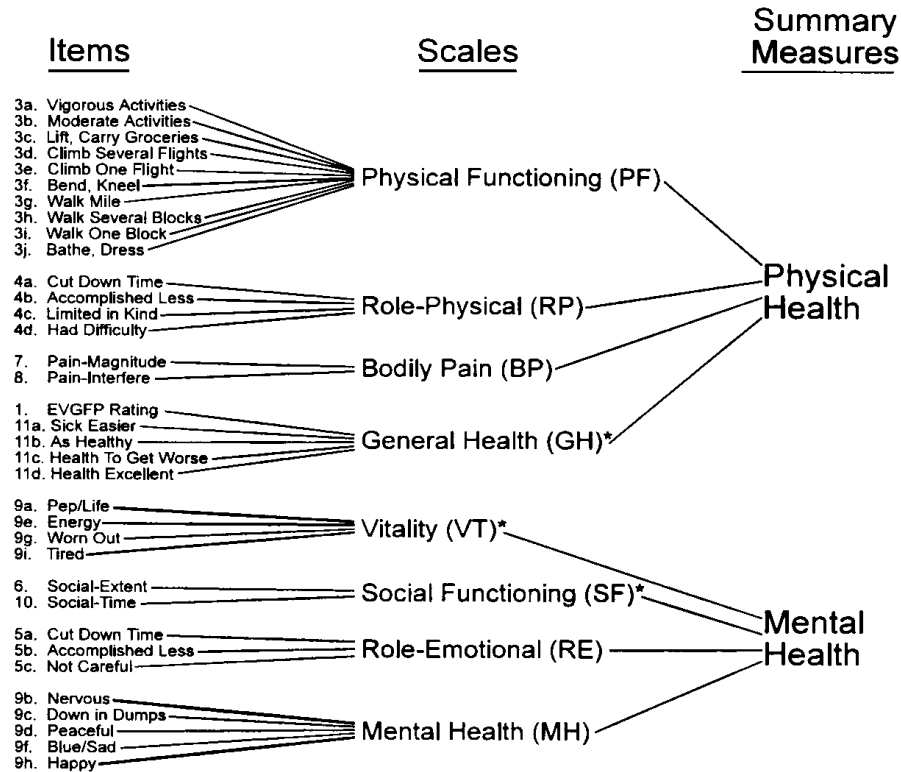


References: Fisher Jr WP, Eubanks RL, Marier RL. Equating the MOS SF36 and the LSU HSI physical functioning scales. *J Outcome Meas.* 1997;1:329–62.

Liegl G, Gandek B, Fischer FH, Bjorner JB, Ware JE, Rose N, Fries JF, and Nolte S. Varying the item format improved the range of measurement in patient-reported outcome measures assessing physical function, *Arthritis Research & Therapy* 2017;19: 66.

Next Step is One Item Per Domain: However, These Items Aren't the Best Measures of Their Respective Domains

FIGURE 3.1 SF-36 MEASUREMENT MODEL



* Significant correlation with other summary measure.

Why Improve Single-item Measures?

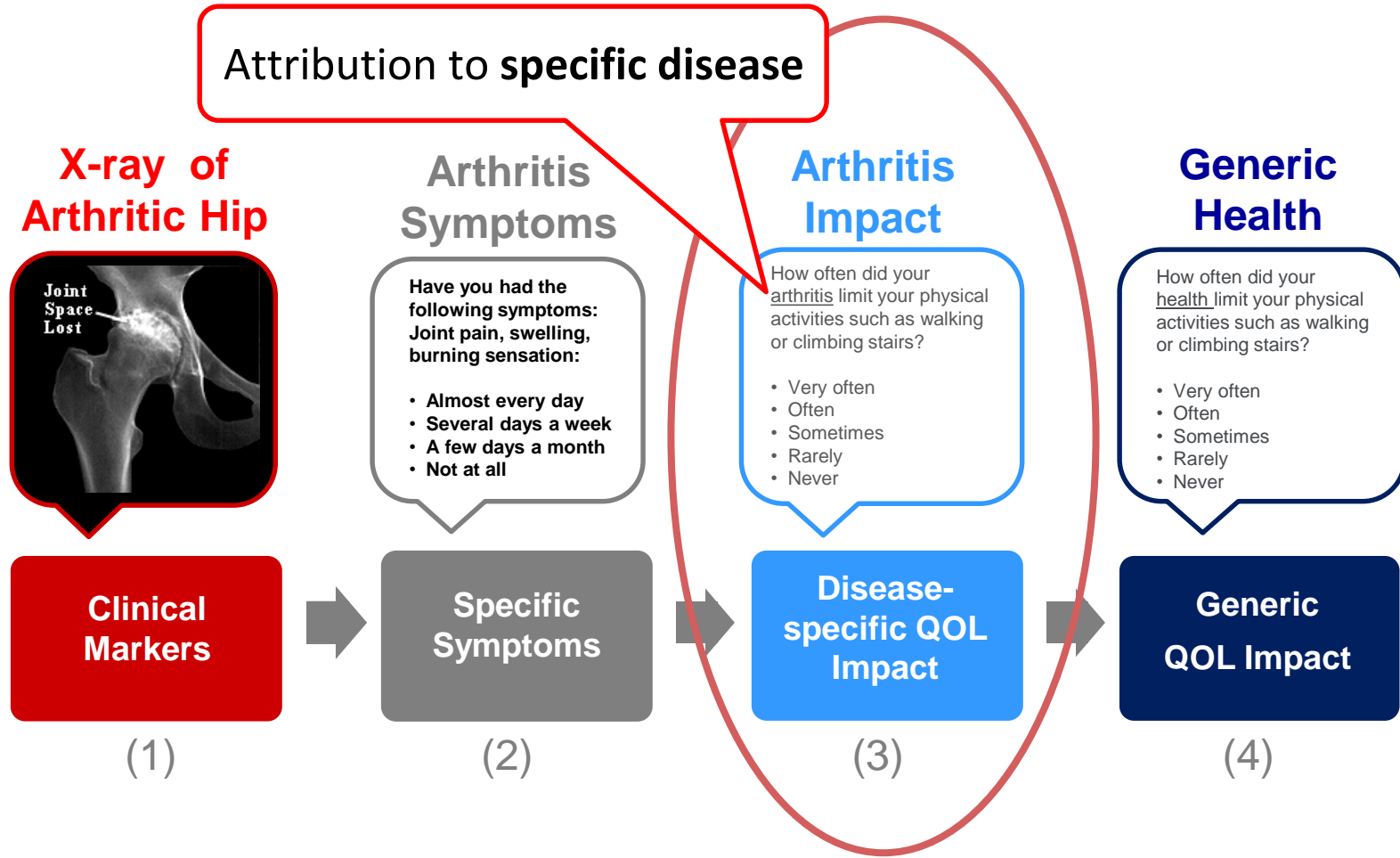
- All surveys begin with the first questionnaire item
- The first item may be the only item per domain in the shortest short-form
- Also, because the first item determines what happens next in adaptive surveys, finding the best item is crucial in improving efficiency
- Systematic comparisons of item characteristics by domain show that the best items are not included in the most widely-used multiple-item scales

Improved Single-item Measures

Improvements over MOS, PROMIS and utility surveys include single-item measures with the following characteristics:

- Broader content representation,
- Response categories that cover a wider range,
- Directly measure a higher-order concept, and
- Better match between the essence of the domain and the operational definition.

New Approach to Standardized Disease-specific QOL Impact Measurement



What is the QOL Disease Impact Scale (QDIS®)

- A suite of disease-specific measures that are standardized across chronic conditions, including:
 - Comprehensive item bank for use in adaptive testing
 - 7-item fixed-length short form (QDIS-7)
 - Global QOL impact item
- Expands disease-specific item content to be on a par with the most comprehensive generic QOL measures
- Uses disease-specific attributions to increase responsiveness to one condition or disease
- All measures scored using the same underlying metric and normed in the US chronically-ill population

Source: Ware JE, Gandek B, Guyer R and Deng N. Standardizing Disease-specific Quality of Life Measures Across Multiple Chronic Conditions: Development and Initial Evaluation of the QOL Disease Impact Scale (QDIS®), *Health and Quality of Life Outcomes*, 2016; 14: 84.

Content and Respondent Burden for General and Disease-Specific Measures

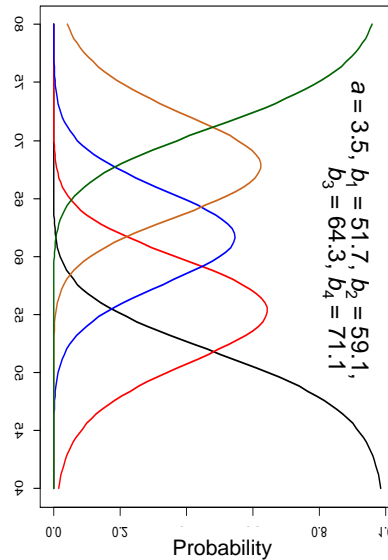
Health Domains*	Comprehensive Generic			Disease-Specific						QOL Disease Impact Scale (QDIS)
	SIP	MOS SF-36	PROMIS 29-57	KD DQOL	ML-QOL	ML-HFQ	SAQ	St Georg	WO-MAC	
Physical functioning	●	●	●	●		●	●	●	●	●
Social functioning	●	●		●	●	●		●		●
Role functioning	●	●	●	●	●	●		●		●
Psychological distress	●	●	●	●	●	●	●			●
Health perceptions		●				●		●		●
Pain (bodily)		●	●	●			●		●	●
Energy/fatigue		●	●		●	●		●		●
Psychological well-being		●								●
Sleep	●		●	●		●		●		●
Cognitive functioning	●									●
Quality of life					●		●			●
Health transition (B/S/W)		●								
Sexual functioning				●	●	●		●		
<u>Specific symptoms</u>				●	●	●	●	●	●	
<u>Specific treatments</u>				●	●	●	●			

SIP=Sickness Impact Profile (1976); **MOS FWBP**=MOS Functioning and Well-Being Profile, 149 items (1992); **PROMIS®**=Patient Reported Outcomes Measurement Information System (2004-on); **DQOL**=Diabetes Quality of Life Measure; **KDQOL**=Kidney Disease Quality of Life Instrument; **MLHFQ**=Minnesota Living with Heart Failure® Questionnaire; **SAQ**=Seattle Angina Questionnaire; **St George**=St George's Respiratory Questionnaire; **WOMAC®**=Western Ontario and McMaster Universities Osteoarthritis Index; **QDIS®**=QOL Disease Impact Scale (Ware et al, 2016). **Source for comparison:** Adapted from Ware 1987; Ware, 1995.

Anatomy of a Survey Item: What Makes it Generic or Disease-specific?

During the past 4 weeks, how often did your **heart problem** limit your ability to do your everyday activities such as work, school or chores?

Very often
Often
Sometimes
Rarely
Never



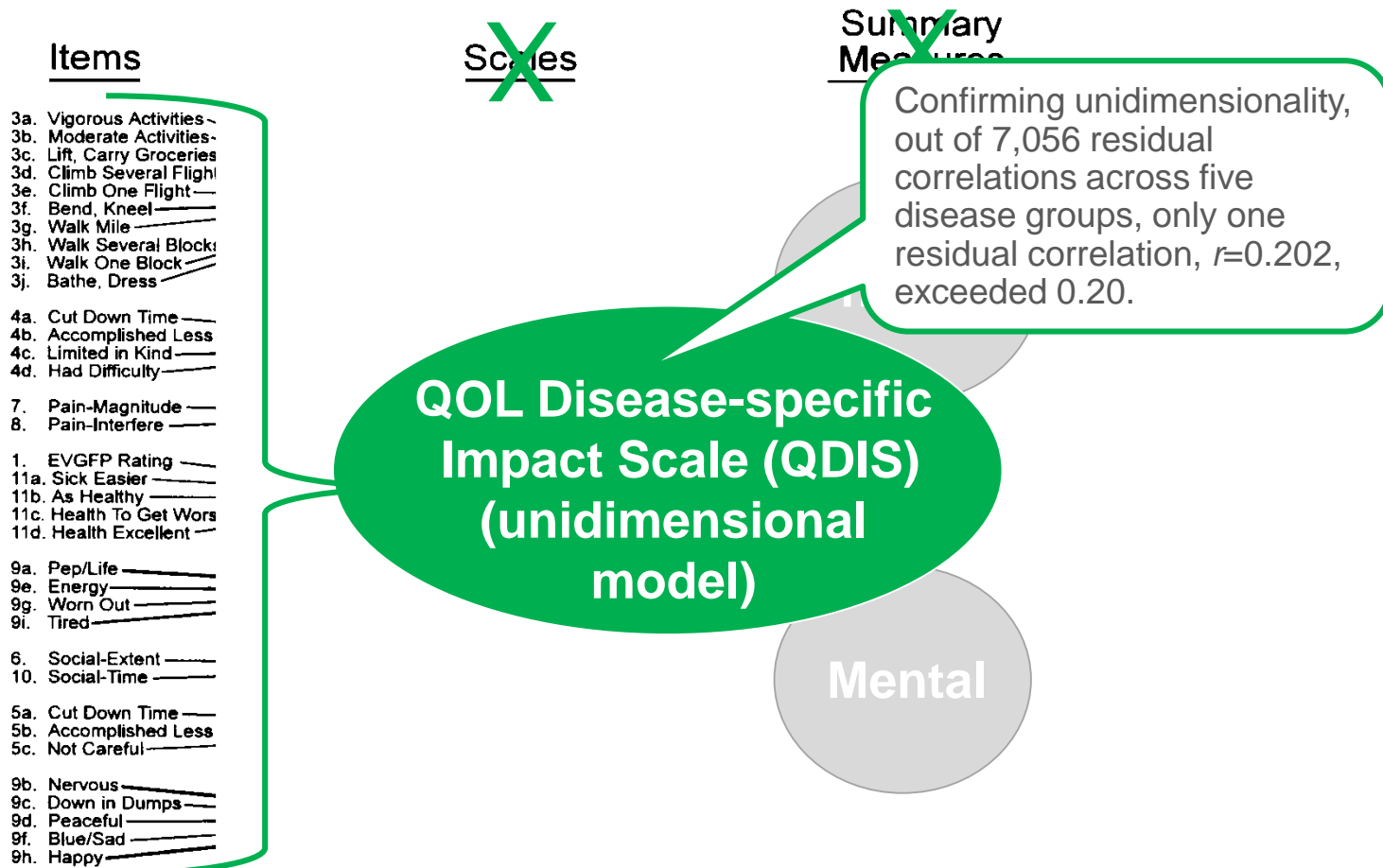
**Disease-specific
attribution**

References:

Ware JE, Gandek B, Guyer R, Deng N. Standardizing disease-specific quality of life measures across multiple chronic conditions: Development and initial evaluation of the QOL Disease Impact Scale (QDIS®). *Health and Quality of Life Outcomes* 2016; 14:84.

Deng N, Anatchkova MD, Waring ME, Han KT, Ware JE. Testing item response theory invariance of the standardized Quality-of-life Disease Impact Scale (QDIS®) in acute coronary syndrome patients: Differential functioning of items and test. *Quality of Life Research* 2015; 24(8): 1809-22.

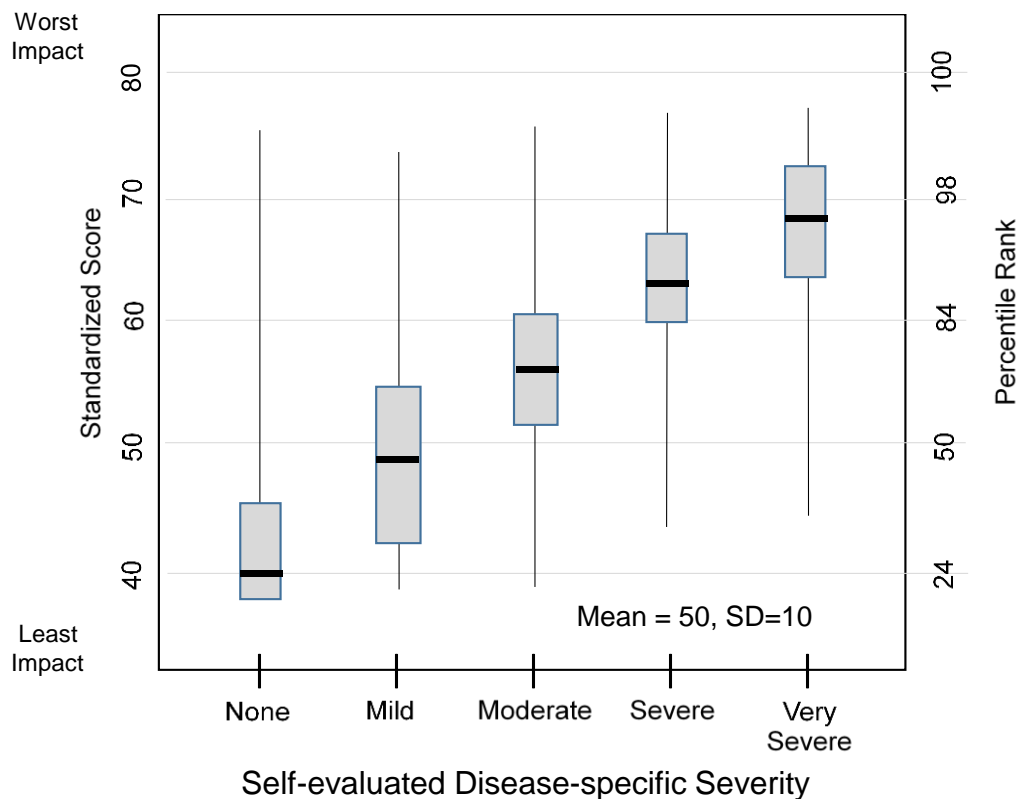
In Contrast to Generic Health, the Model for Disease-Specific Attributions is 1-Factor



References:

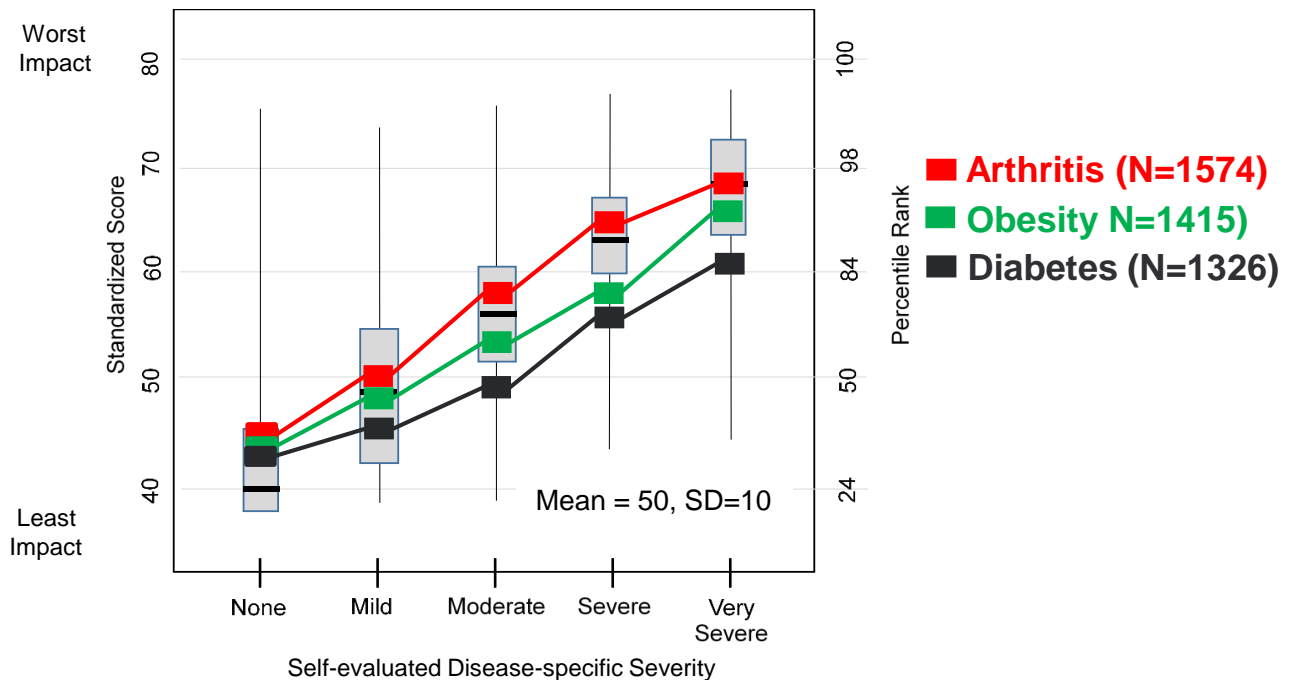
Ware JE, Gandek B, Guyer R, Deng N. Standardizing disease-specific quality of life measures across multiple chronic conditions: Development and initial evaluation of the QOL Disease Impact Scale (QDIS®). *Health and Quality of Life Outcomes* 2016; 14:84.

QDIS Scores Increase with Disease Severity, US General Chronically-ill Population (N=5351)



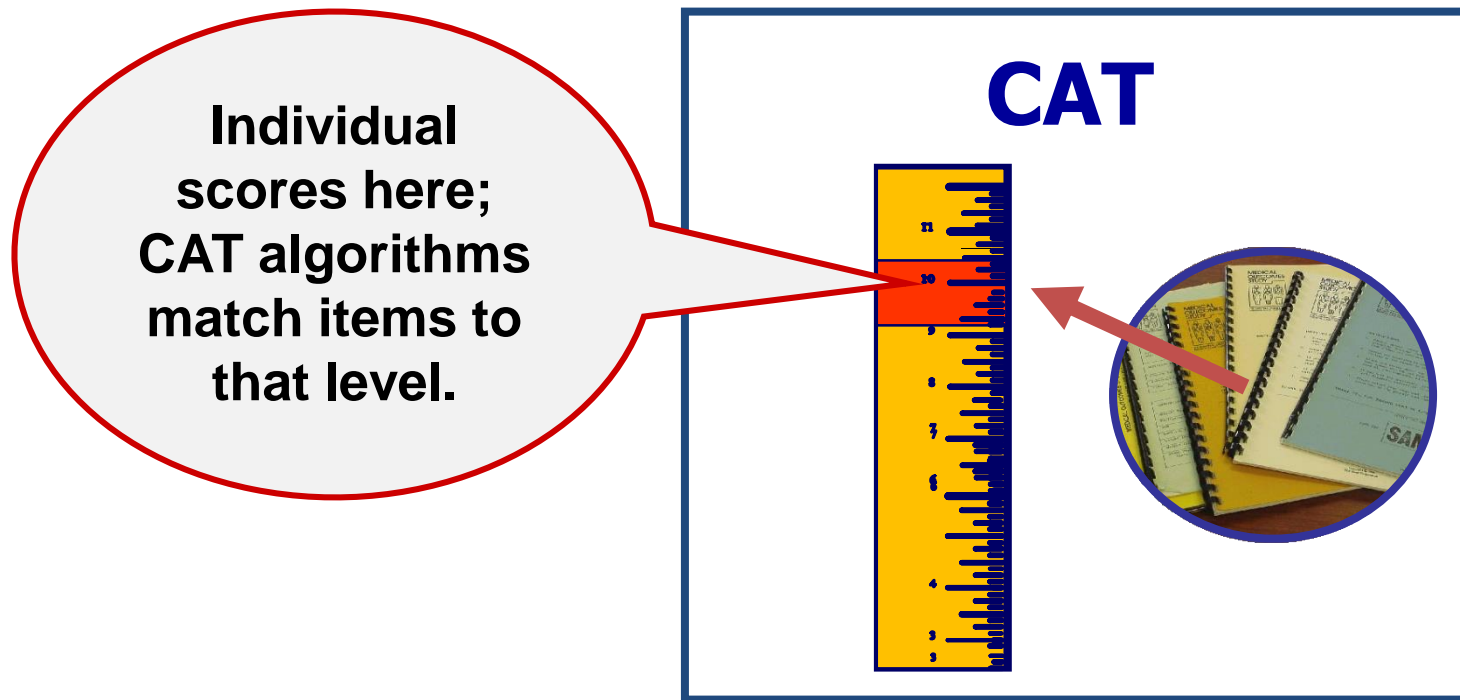
Source: Ware JE, Gandek B, Guyer R and Deng N. Standardizing Disease-specific Quality of Life Measures Across Multiple Chronic Conditions: Development and Initial Evaluation of the QOL Disease Impact Scale (QDIS®), *Health and Quality of Life Outcomes*, 2016; 14: 84.

Plot of QDIS Scores Across Groups Differing in Disease Severity, US General Chronically Ill Population (N=5351)



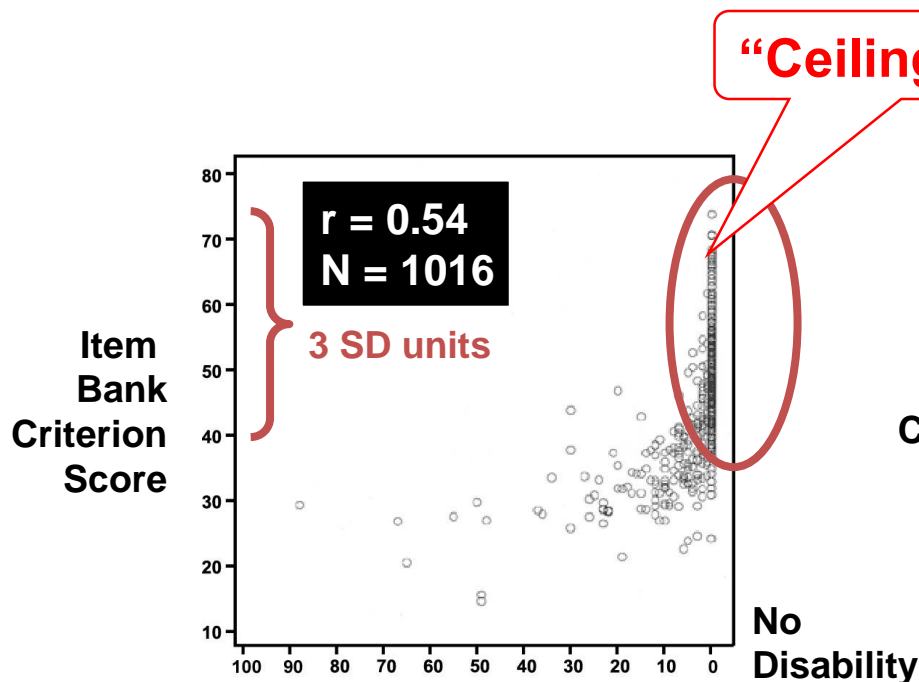
Source: Ware JE, Gandek B, Guyer R and Deng N. Standardizing Disease-specific Quality of Life Measures Across Multiple Chronic Conditions: Development and Initial Evaluation of the QOL Disease Impact Scale (QDIS®), *Health and Quality of Life Outcomes*, 2016; 14: 84.

Solution to the Practical vs Precise Measurement Problem: Computerized Adaptive Test (CAT) Assessments

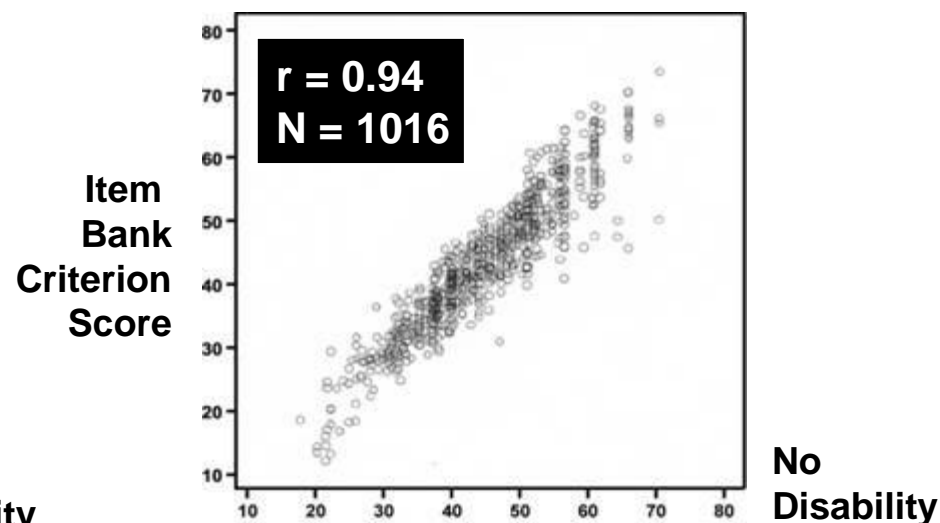


CAT = Computerized Adaptive Testing

Practical Implications of CAT-Based Assessment Demonstrated in 2000



Static 5-Item Headache Pain Measure

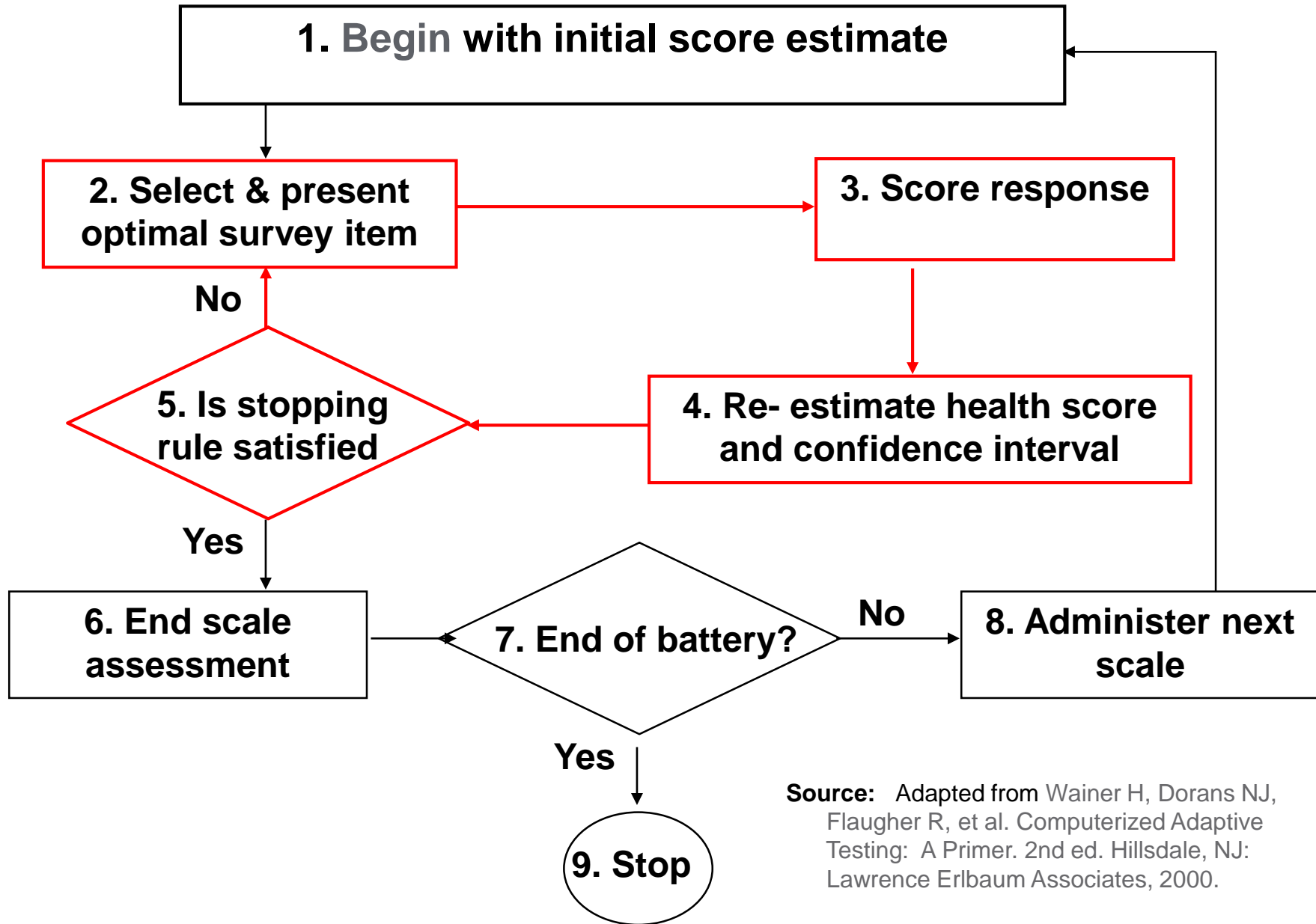


Adaptive 5-Item Headache Pain Measure

References:

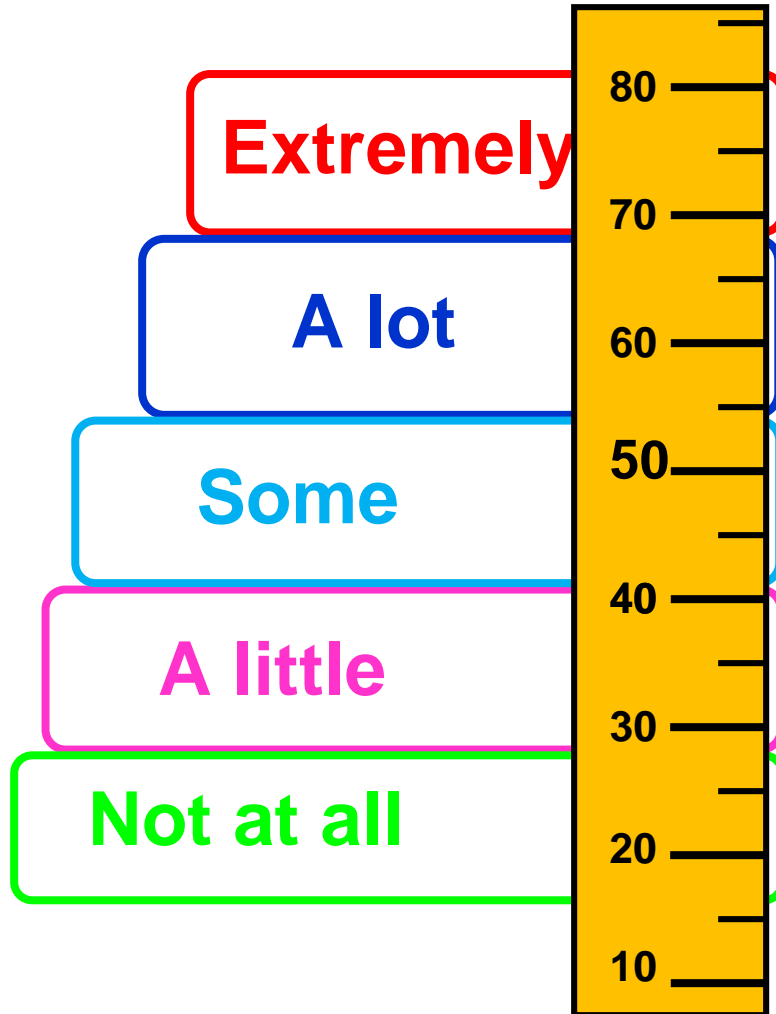
- Ware JE, Jr., Bjorner JB, Kosinski M: Practical implications of item response theory and computerized adaptive testing: A brief summary of ongoing studies of widely used headache impact scales. *Medical Care* 2000;38:1173-82.
- Ware JE, Jr., Kosinski M, Bjorner JB, et al. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Quality of Life Research* 2003;12(8):935-52.

Logic of Computerized Adaptive Testing (CAT)

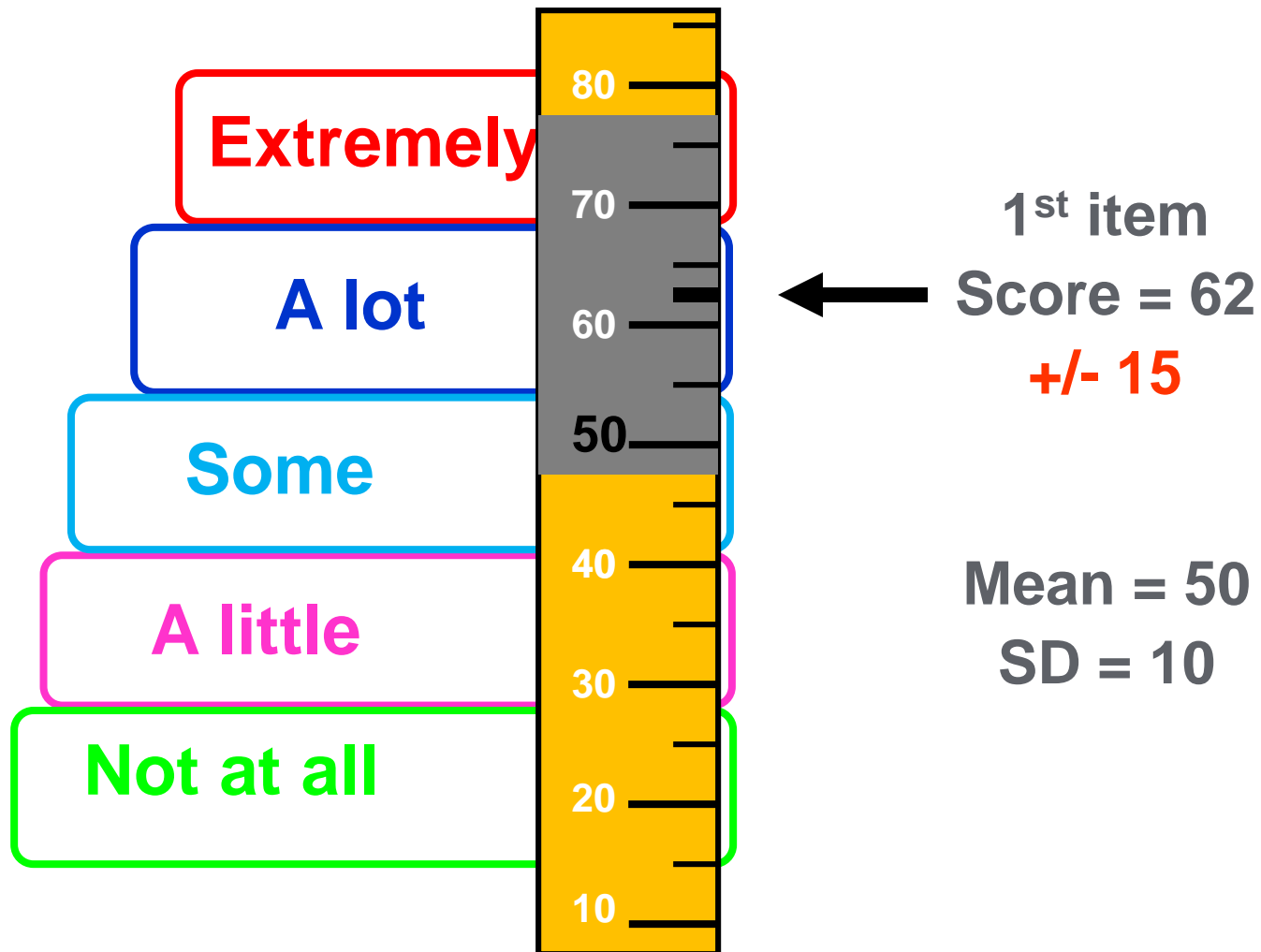


Source: Adapted from Wainer H, Dorans NJ, Flaugher R, et al. Computerized Adaptive Testing: A Primer. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 2000.

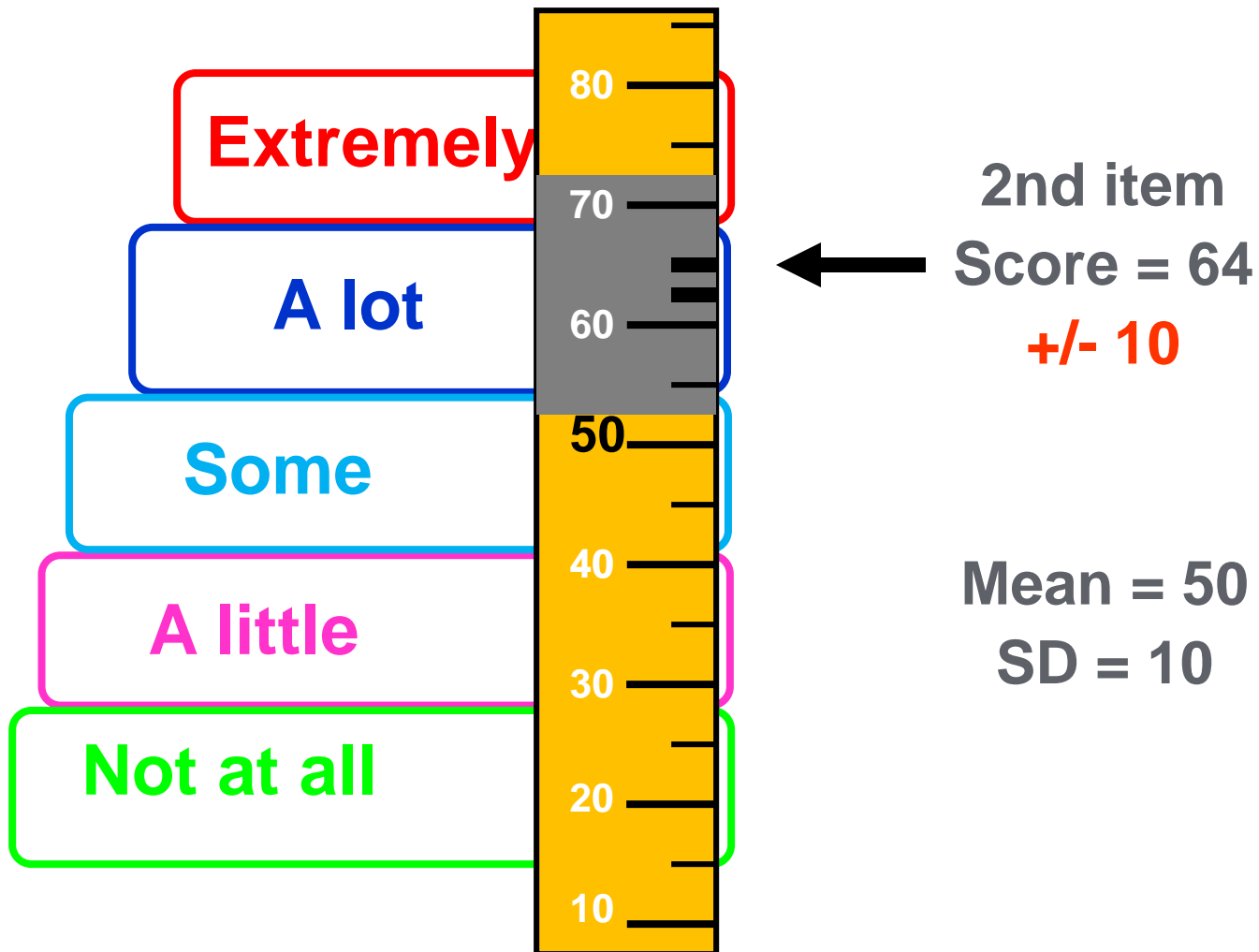
Adaptive Software Can Substantially Improve Measurement Efficiency



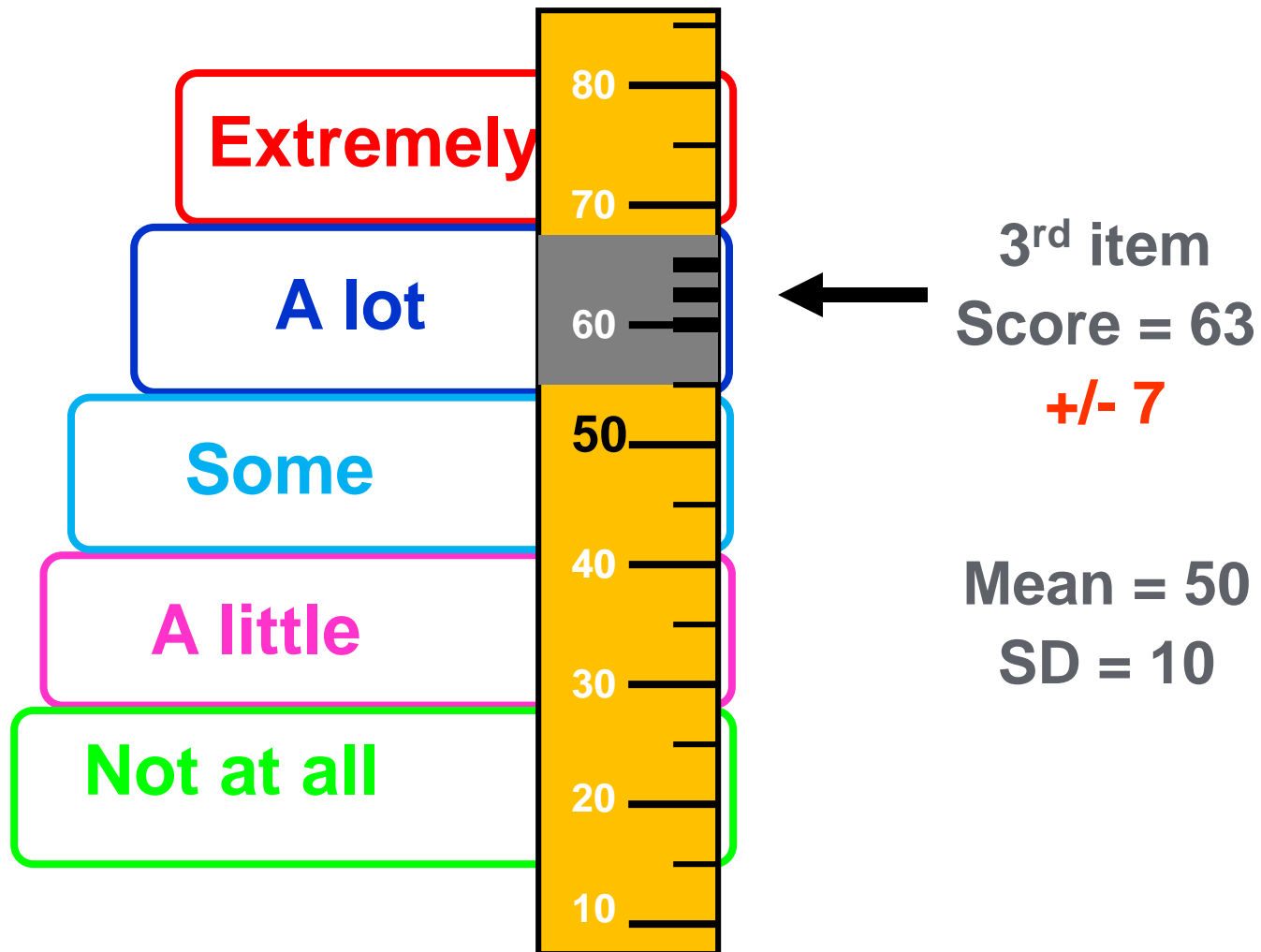
1st Item Yields Noisy Score Estimate



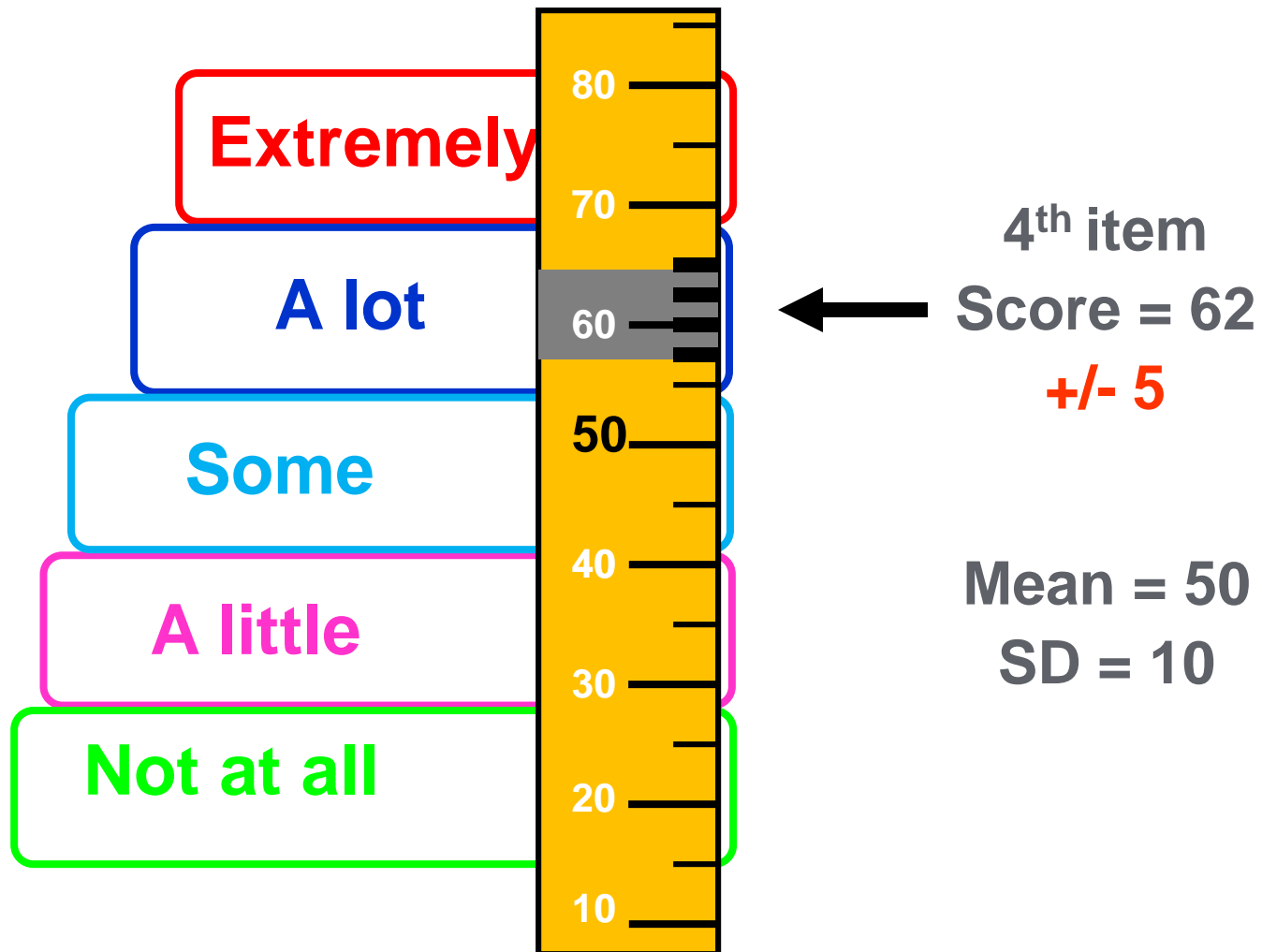
2nd Item Reduces CI by 1/3



3rd Item Cuts CI by 1/2



4th Item Cuts CI by 2/3



Advantages of Computerized Adaptive Test (CAT) A

- More accurate individual risk scores and outcomes monitoring
- Brevity of a short form – 70%+ reduction in response burden
- Elimination of “ceiling” effects, with improved items
- Can be administered using various electronic data collection technologies
- Monitor data quality in real time

However, because of the number of generic domains and comorbid conditions, routine CAT's won't solve the respondent burden problem.

Alternate Forms of QDIS

- 1 global impact item
- 7-item static form
- MCC form
- CAT

In the past 4 weeks, how much did your asthma limit your everyday activities or your quality of life?

Not at all



1

A little



2

Some



3

A lot



4

Extremely



5

Copyright © 2010-2017 John Ware Research Group, Inc. All rights reserved.

Alternate Forms of QDIS

- 1 global impact item
- 7-item static form
- MCC form
- CAT

Quality of Life Impact Survey for Asthma

1. In the past 4 weeks, how much did your asthma limit your everyday activities or your quality of life?

Not at all	A little	Some	A lot	Extremely
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. In the past 4 weeks, how often...

	Never	Rarely	Sometimes	Often	Very often
• Integer vitae quam a nial dapibus tincidunt vel ac diam?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Duis fermentum diam in metus lacina blandit volutpat ipsum scelerisque?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Pellentesque erat consectetur tellus placerat sed faucibus nulla pulvinar?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Maecenas bibendum dui semper augue placerat conorndo ultrices metus efficitur?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Lorem ipsum dolor sit amet consectetur adipiscing elit?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Nunc consectetur purus in massa aliquam pretium?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Copyright © 2010-2017 John Ware Research Group, Inc. All rights reserved.

Alternate Forms of QDIS

- 1 global impact item
- 7-item static form
- MCC form
- CAT

Chronic Conditions and How They Affect You

Have you ever been told by a doctor or other health professional that you had any of the following conditions? On the left, please answer NO or YES for each condition. If YES, on the right, please answer the question for each condition you have.

Please answer NO or YES for each condition	If YES		If YES, in the past 4 weeks, how much did this CONDITION limit your everyday activities or your quality of life?				
	NO	YES	Not at all	A little	Some	A lot	Extremely
Congestive heart failure or enlarged heart	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diabetes or high blood sugar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stroke	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cancer, except skin cancer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

QDIS® in Chronic Condition Checklist format. Copyright © 2010-2017 John Ware Research Group, Inc. All rights reserved.

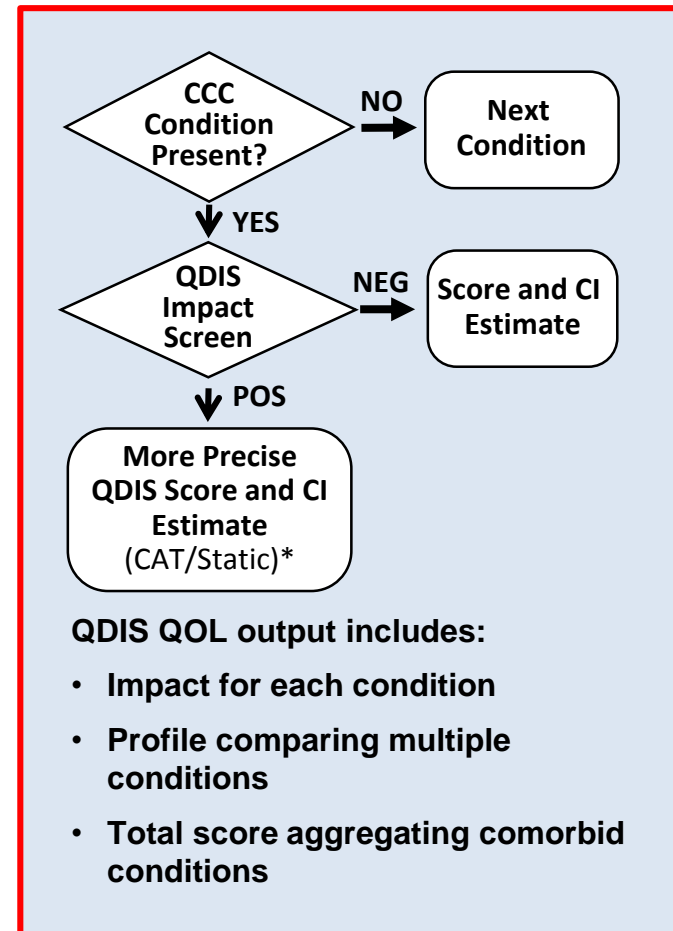
MCC: Multiple chronic conditions

QDIS items © 2010-2017 John Ware Research Group Inc. All rights reserved.

Alternate Forms of QDIS

- 1 global impact item
- 7-item static form
- MCC form
- **CAT**

Chronic Condition Checklist (CCC) & QDIS



Convergent-Discriminant Validity Tests Require Multiple Constructs and Multiple Methods

- **Convergent validity** is supported when substantial correlations are observed between different methods of measuring the same construct (e.g., disease, health domain)
- **Discriminant validity** is supported when much lower correlations are observed between different constructs measured by the same method

Reference: Campbell DT and Fiske DW, Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959;56(2):81-105.

Convergent-Discriminant Validity Correlation Matrix, Comorbid Arthritis and Obesity (N=337)

Conditions & Methods	Arthritis			Obesity		
	Ar1	Ar2	Ar3	Ob1	Ob2	Ob3
Arthritis (Ar)						
Ar1 Sx – Joint pain						
Ar2 Severity	.51					
Ar3 QOL impact	.49	.75				
Obesity (Ob)						
Ob1 Sx – BMI	.05	.15	.21			
Ob2 Severity	.14	.22	.25	.55		
Ob3 QOL impact	.19	.24	.32	.39	.60	

Convergent validity: Substantial correlations between different methods of measuring the same disease: $r = .49$ to $.75$

Discriminant validity: Lower correlations between different diseases measured with the same ($r = .05$ to $.32$) and different methods ($r = .14$ to $.25$)

Convergent validity: $r = .39$ to $.60$

Note: Body mass index (BMI)

Source: Ware JE, Gandek B, Allison J., The Validity of Disease-specific Quality of Life Attributions Among Adults with Multiple Chronic Conditions. *International Journal of Statistics in Medical Research* 2016; 5: 1-24.

Responsiveness of Disease-specific and Generic QOL Measures Across Groups Reporting Better and Worse 9-month Outcomes

Measures	Much Better (n=244)	Somewhat Better (n=245)	Same (n=1181)	Somewhat Worse (n=282)	Much Worse (n=378)	F-ratio	RV ^a
Disease-specific (-)							
QDIS-7	-2.76	-0.04	1.29	3.20	5.87	29.8*	1.00**
General (+)							
SF-8 Physical (PCS)	1.36	-0.71	-0.28	-3.22	-4.98	14.2*	0.47
SF-8 Mental (MCS)	1.33	-0.47	0.14	-0.33	-0.91	2.1	0.07

Note: For each outcome, self-evaluated change groups were defined in response to the question: "Compared to nine months ago, how much better or worse is your <DISEASE> now?", where DISEASE was a pre-defined and confirmed for nine chronic conditions analyzed here in the aggregate (total N=2330).

^aRelative validity (RV) is the ratio of the comparator general health F-statistic over the best (QDIS) F-statistic

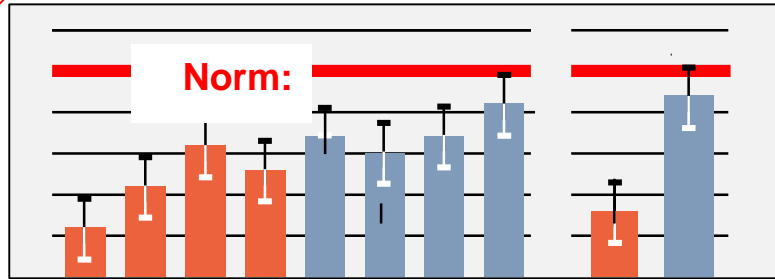
*Significant F-ratio for comparison of average changes

**RV significantly greater in comparison with general measures using bootstrap method.

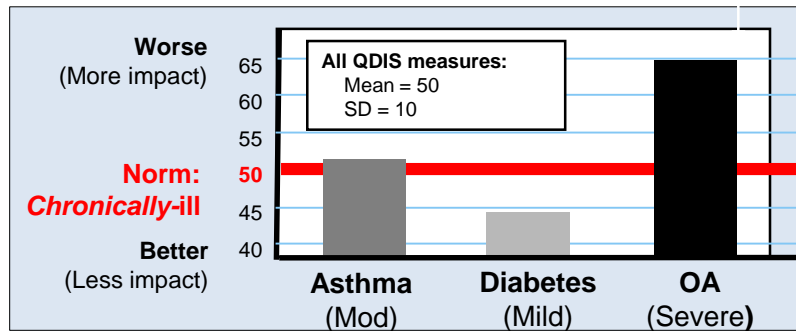
Source: Ware JE, Gandek B, Guyer R and Deng N. Standardizing Disease-specific Quality of Life Measures Across Multiple Chronic Conditions: Development and Initial Evaluation of the QOL Disease Impact Scale (QDIS®), Health and Quality of Life Outcomes, 2016; 14:84.

Summary

- ✓ Comparisons of outcomes across applications requires QOL measurement standardization
- ✓ To make results more actionable, generic and disease-specific QOL must be displayed on the same dashboard



SF-36 Profile & Summaries



QDIS Multiple Chronic Condition Profile

practical, a new generation of single-

s and utility items for common
single-items with:

otive content

wider range

der concept

f domain and operational definition

asurement system that monitors QOL

n when necessary, and automatically

erity of multiple chronic conditions is

Summary

- ✓ • Comparisons of effectiveness across applications requires QOL measurement standardization
- ✓ • To make results more actionable, generic and disease-specific QOL must be displayed on the same dashboard
- ✓ • To make data collection more practical, a new generation of single-item super-short-forms is required
- ✓ • Improvements over MOS, PROMIS and utility items for common domains have been linked to single-items with:
 - Broader representation of descriptive content
 - Response categories covering a wider range
 - Direct measurement of higher-order concept
 - Better match between essence of domain and operational definition
- ✓ • A more aggressive adaptive measurement system that monitors QOL with “super SF items,” drills down when necessary, and automatically adapts to the presence and severity of multiple chronic conditions is required

