Basic Methods for Measurement of Patient-Reported Outcome Measures

Ron D. Hays, Ph.D. (drhays@ucla.edu)
UCLA/RAND

ISOQOL Conference on Patient Reported Outcomes in Clinical Practice, Workshop #4
Hilton Budapest (June 24, 2007, 1-5pm)
http://www.gim.med.ucla.edu/FacultyPages/Hays/

Four Hours

I. Conceptualization of PRMs and Measurement of PROs
II. Evaluating Individual Change
III. Psychometric Properties of a Good Measure
IV. Steps to Develop a Measure and Residual Questions

{10 minute breaks at about 2, 3, & 4pm}

3 Paracute/4 People Dilemma

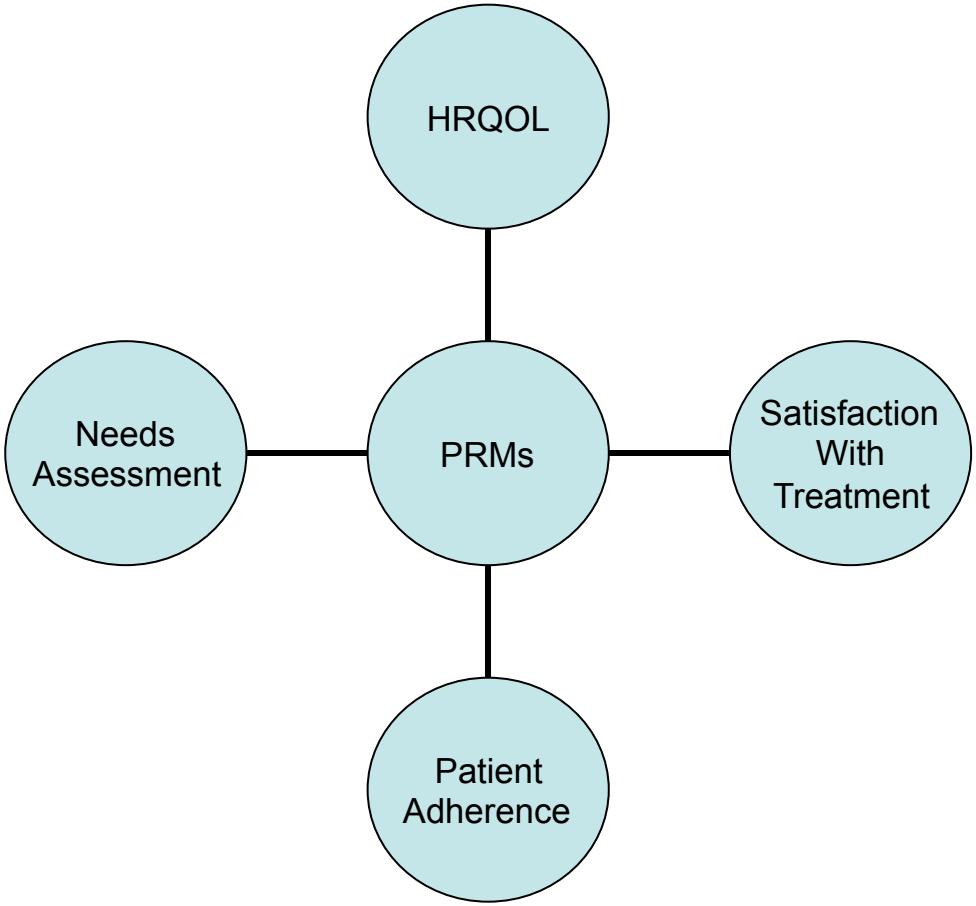
- World's Smartest Man
- George Bush
- Pope
- Boy Scout

Part I: Conceptualization of PRMS and Measurement of PROs

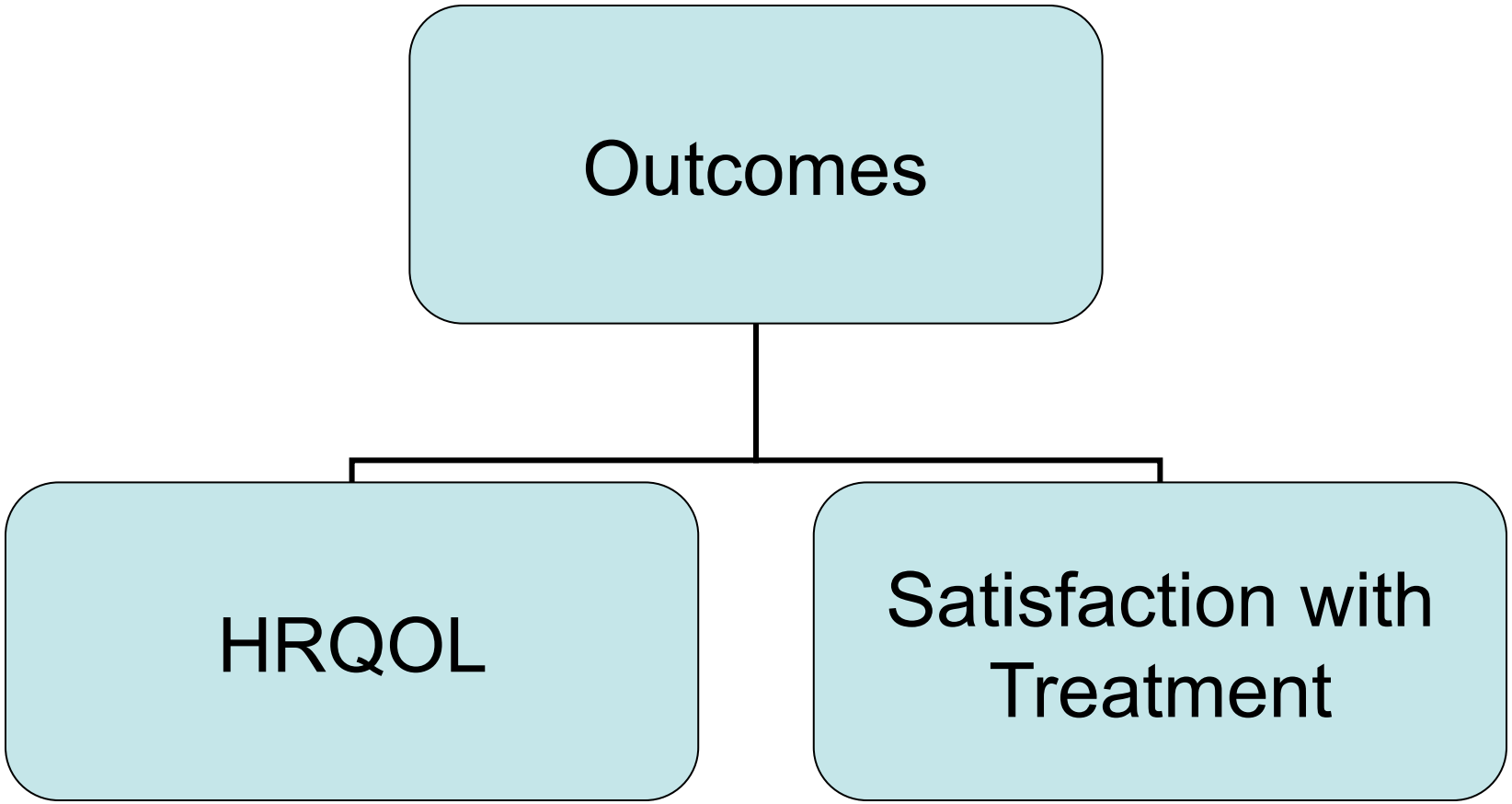
Patient-Centered Care

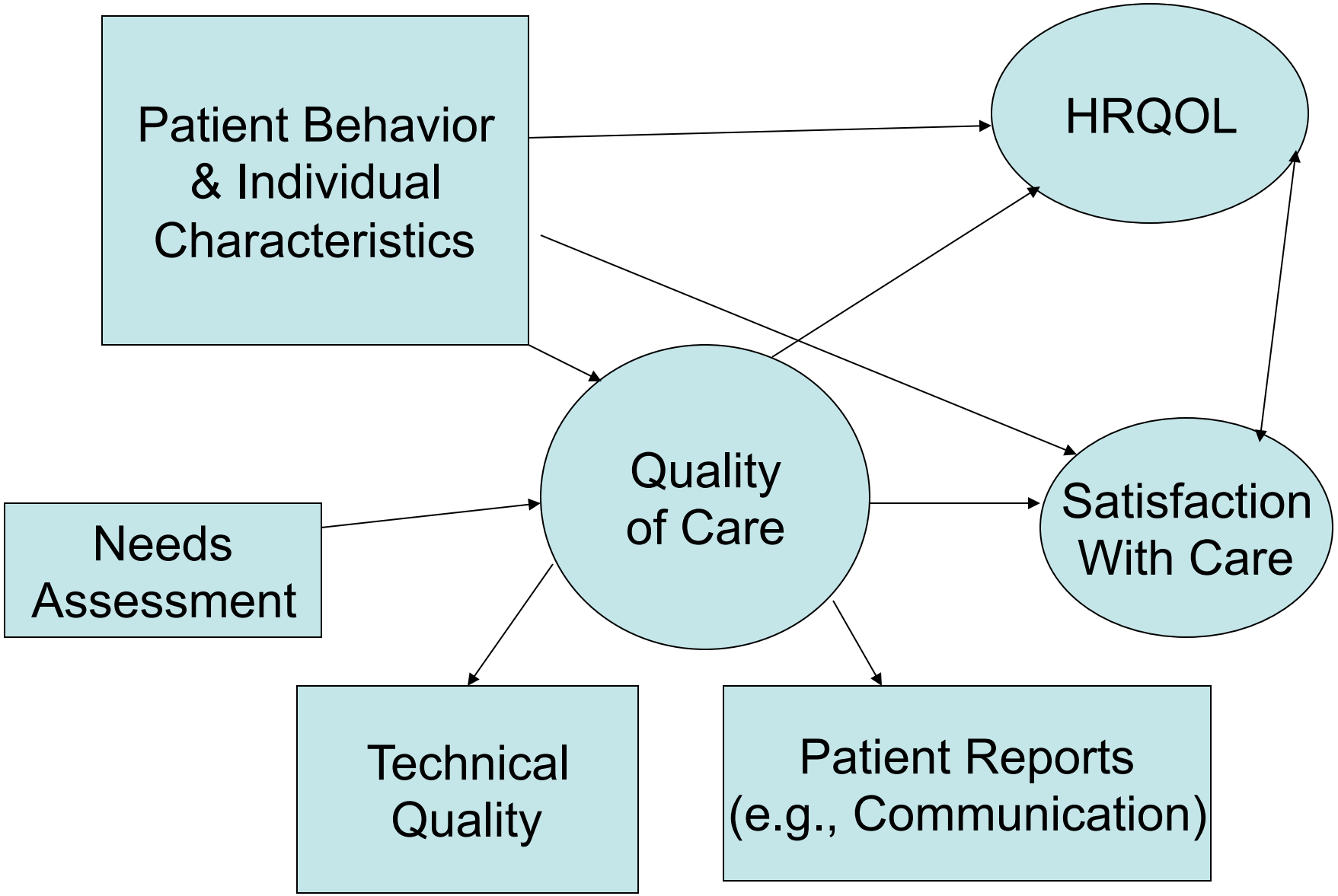
- “Respectful and responsive to individual patient preferences, needs, and values” (IOM, 2001, p. 6)
- Patient-centered care requires knowing what is important to patient (needs)
- Extent to which providers are meeting the needs of their patients is seen in
  - Patient evaluations of care
  - Health-related quality of life

Patient-Reported Measures (PRMs)



Outcomes of Care




Patient Behavior & Individual Characteristics
HRQOL
Quality of Care
Satisfaction With Care
Needs Assessment
Technical Quality
Patient Reports (e.g., Communication)

Kingston Needs Assessment Questionnaire--Cancer

- 52 items
- Need domains
  - Symptoms control
  - Information
  - Support services
  - Experience at cancer center
  - Coordination of care

Needs Assessment Items

- Information about possible treatment options
- Information about possible benefits of the treatments
- Information about cancer and my specific case
- Information about possible harms (side effects) of the treatment
- Information about treatment procedures
- If you were seen in the ED, being seen in a reasonable length of time.
- If you were hospitalized for cancer-related care, being admitted to the hospital in a reasonable length of time.
- Good communication among all health professionals involved in your case (at the cancer center, in hospital, in the community)
- Receiving test results in a reasonable length of time
- Pain management
- Getting your first appointment within a reasonable length of time.

Take-away Conceptual Points

- PROs (Satisfaction with treatment, HRQOL) are:
  - Subset of PRMs
    - U.S. FDA (2006) definition of PRO is too broad: "any report coming from patients about a health condition and its treatment"
  - Bottom-line of whether care produces outcomes that is valued by the patient
- Other PRMs tell us what patients
  - want (needs assessment)
  - receive (patient reports about care)
  - do (patient adherence and other health behaviors)
  - have as resources (e.g., social support)

Good PRO Measures

- Reliable
- Valid
- Used by clinician
  - Provides useful information about the patient
    - Knowing that a person is able to get out of bed is probably not informative for an ambulatory patient that the doctor has observed walking briskly into the office.
- Practical for patient to complete

Evaluations of Care

- Reports about care domains
  - Communication
  - Coordination of care
  - Get needed care and get care promptly
- Global ratings of care (Satisfaction)
  - Satisfied versus dissatisfied
  - Excellent to Poor
  - 0-10 Ratings

Reports about Care--Communication

- In the last 12 months, how often did your doctor
  - explain things in a way that was easy to understand?
  - listen carefully to you?
  - show respect for what you had to say?
  - spend enough time with you?

Never, Sometimes, Usually, Always

Global Rating Item

Using any number from 0 to 10 where 0 is the worst health care possible and 10 is the best health care possible, what number would you use to rate all your health care in the last 12 months?

- ☐ 0 WORST HEALTH CARE POSSIBLE
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10 BEST HEALTH CARE POSSIBLE

HRQOL

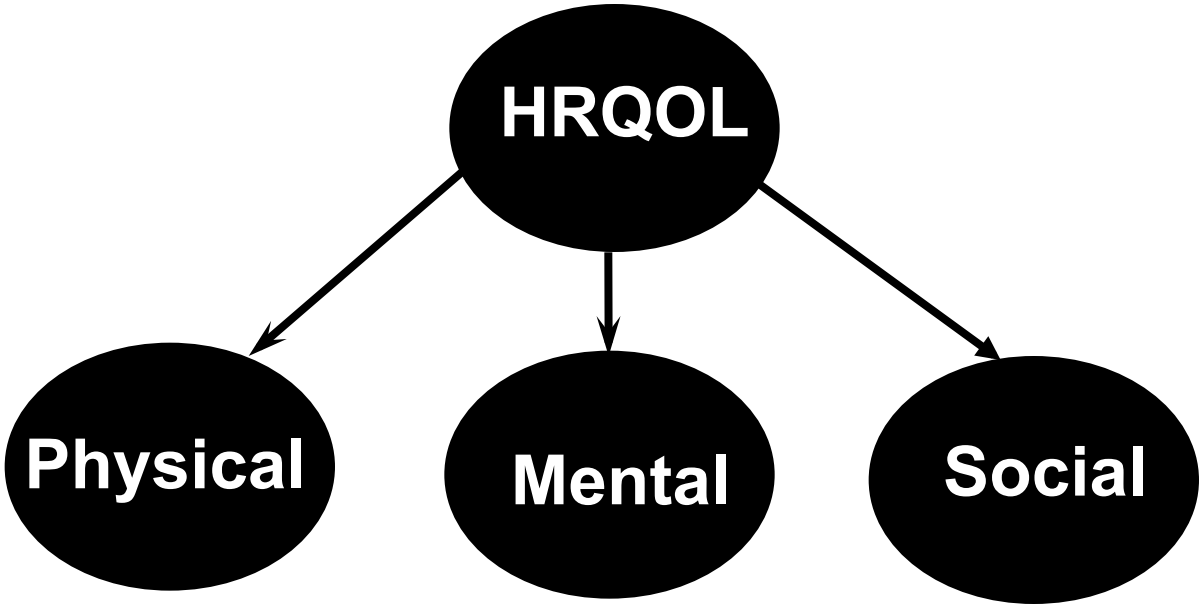
- Domain profiles
  - Physical functioning
  - Social/role participation
  - Pain
  - Energy/fatigue
  - Emotional distress/well-being

- Summary measures
  - Global ratings or general health perceptions
    - In general how would you rate your health?
  - Preference measures

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

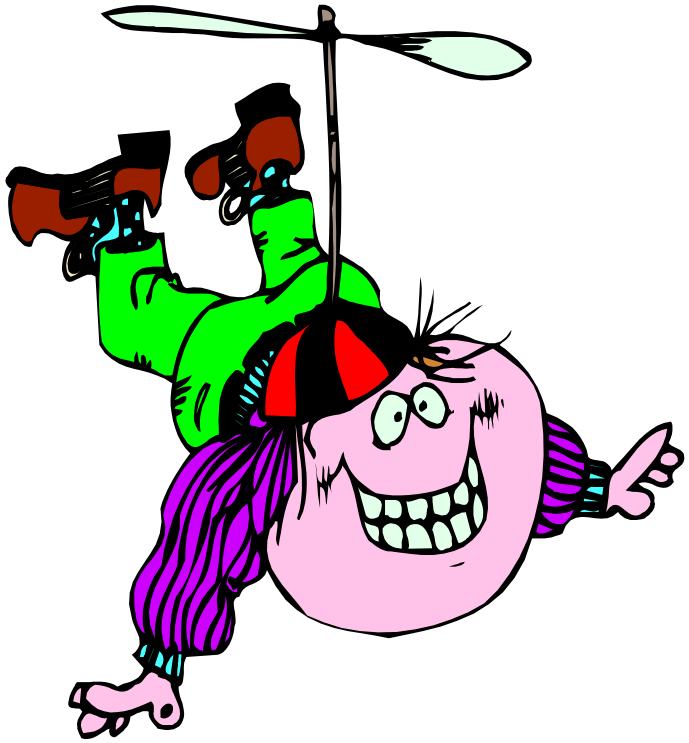
HRQOL is Multi-Dimensional



HRQOL is Not

- Quality of environment
- Type of housing
- Level of income
- Social Support

Types of HRQOL Measures



Profile: Generic vs. Targeted

Preference Measure

SF-36 Generic Profile Measure

- Physical functioning (10 items)
- Role limitations/physical (4 items)
- Role limitations/emotional (3 items)
- Social functioning (2 items)
- Emotional well-being (5 items)
- Energy/fatigue (4 items)
- Pain (2 items)
- General health perceptions (5 items)

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

(If so, how much?)

No, not limited at all

Yes, limited a little

Yes, limited a lot

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

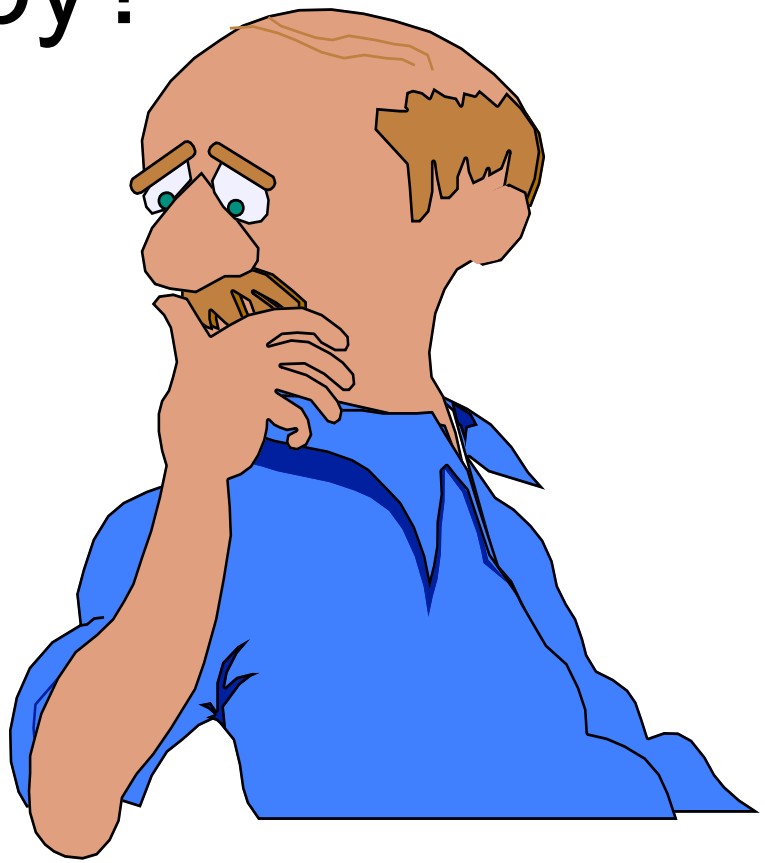
None of the time

A little of the time

Some of the time

Most of the time

All of the time



Scoring HRQOL Scales

- Average or sum all items in the same scale.

- 0 (worst) to 100 (best) possible range (linear) transformation
- Set mean and SD
  - z-score (mean = 0, SD = 1)
  - T-score (mean = 50, SD = 10)

Formula for Transforming Scores

$$X = \frac{(\text{original score} - \text{minimum}) * 100}{(\text{maximum} - \text{minimum})}$$

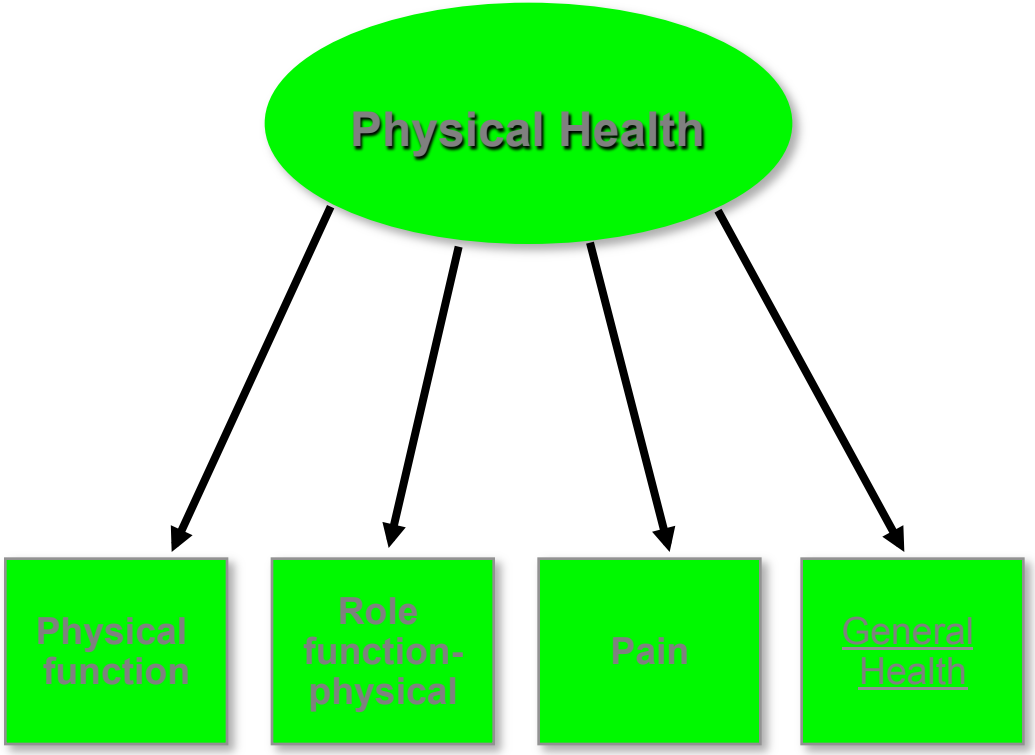
$$Y = \text{target mean} + (\text{target SD} * Zx)$$

$$Z_X = \frac{(X - \bar{X})}{SD_x}$$

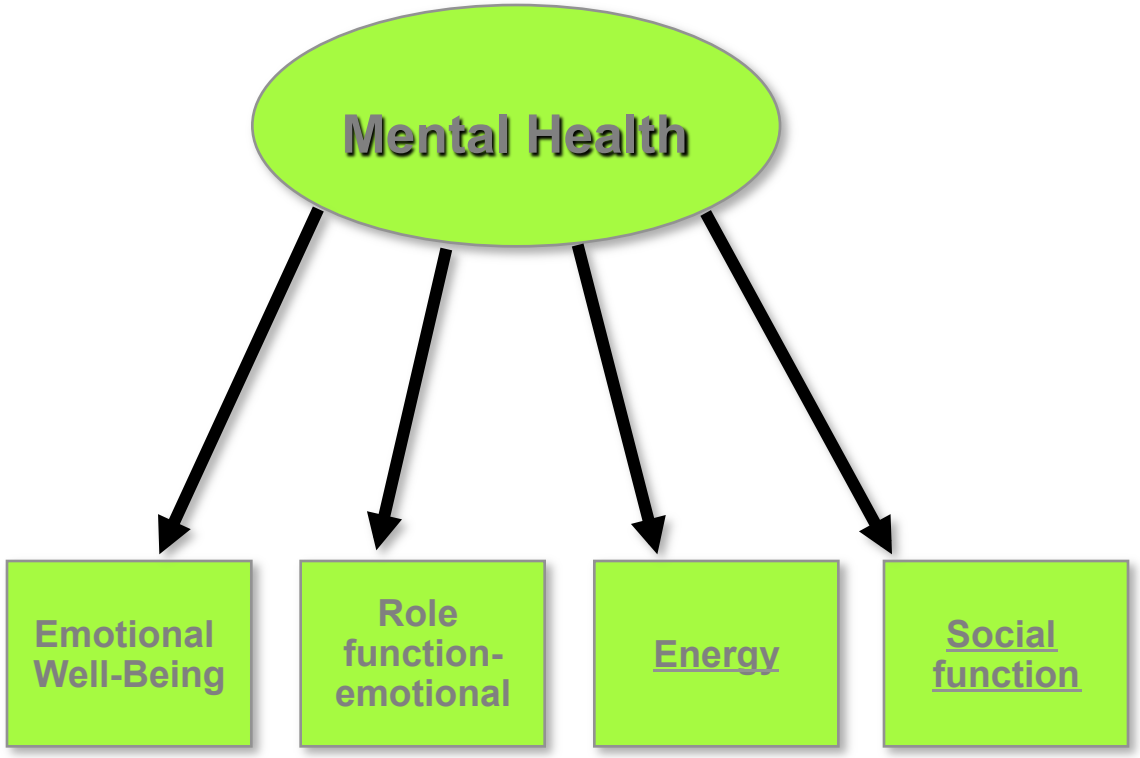
Transforming Scores

1	0	39.5285	-1.26491	37.3509
2	25	39.5285	-0.63246	43.6754
3	50	39.5285	0	50
4	75	39.5285	0.63246	56.3246
5	100	39.5285	1.26491	62.6491
raw	x	SDx	Zx	T-score

SF-36 Physical Health



SF-36 Mental Health



SF-36 Physical Health (PCS) & Mental Health (MCS) Summary Scores

PCS = (PF_Z * .42402) + (RP_Z * .35119) +
(BP_Z * .31754) + (GH_Z * .24954) +
(EF_Z * .02877) + (SF_Z * -.00753) +
(RE_Z * -.19206) + (EW_Z * -.22069)

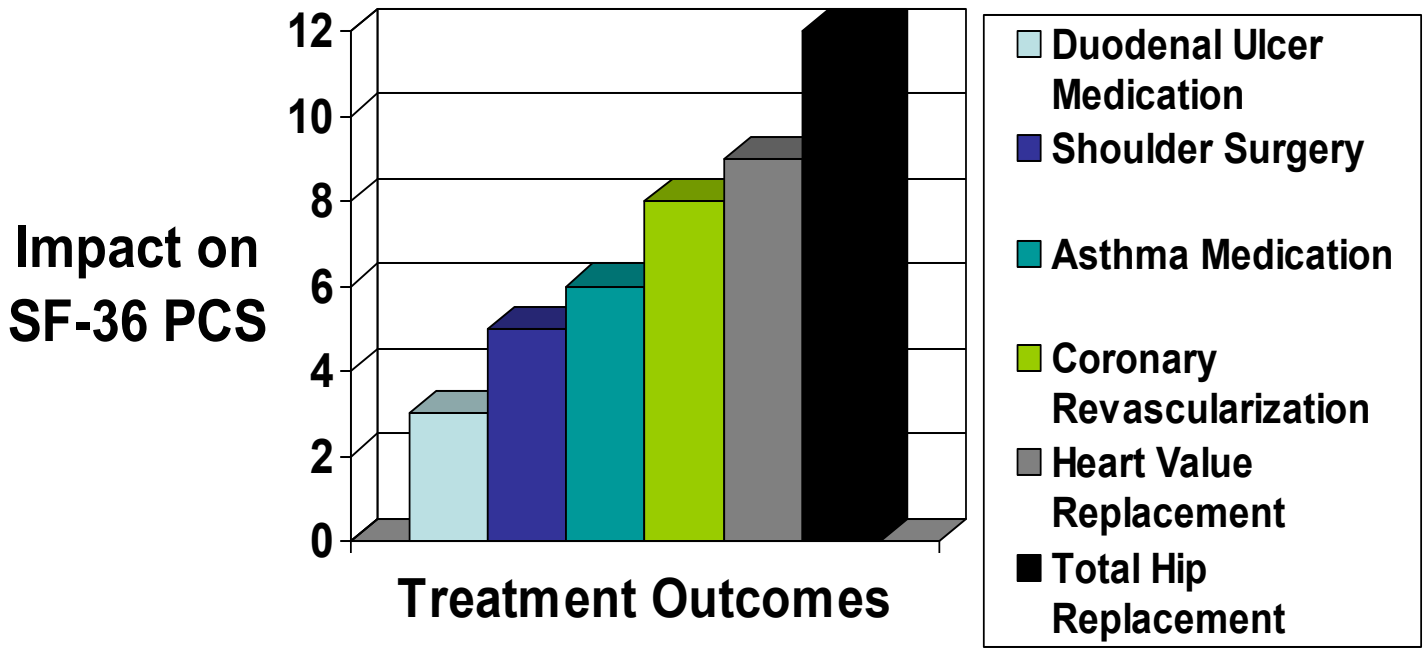
MCS = (PF_Z * -.22999) + (RP_Z * -.12329) +
(BP_Z * -.09731) + (GH_Z * -.01571) +
(EF_Z * .23534) + (SF_Z * .26876) +
(RE_Z * .43407) + (EW_Z * .48581)

T-score Transformation

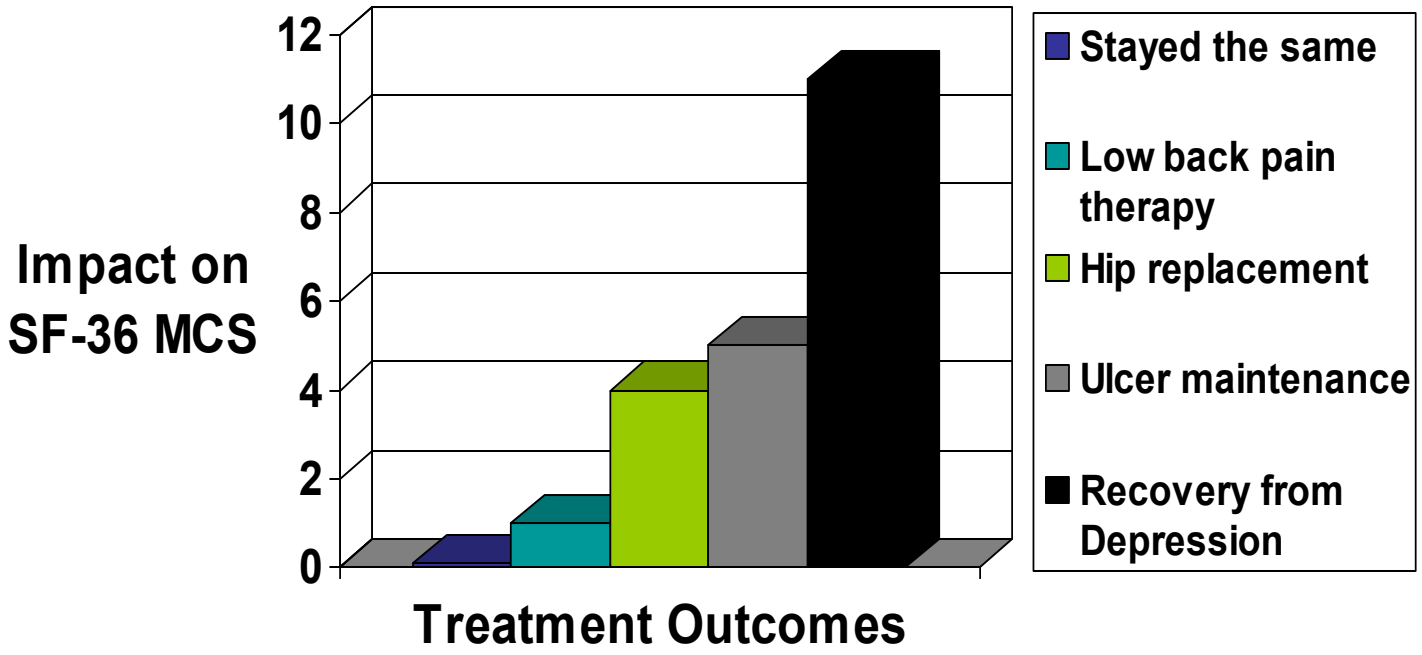
PCS = (PCS_z*10) + 50

MCS = (MCS_z*10) + 50

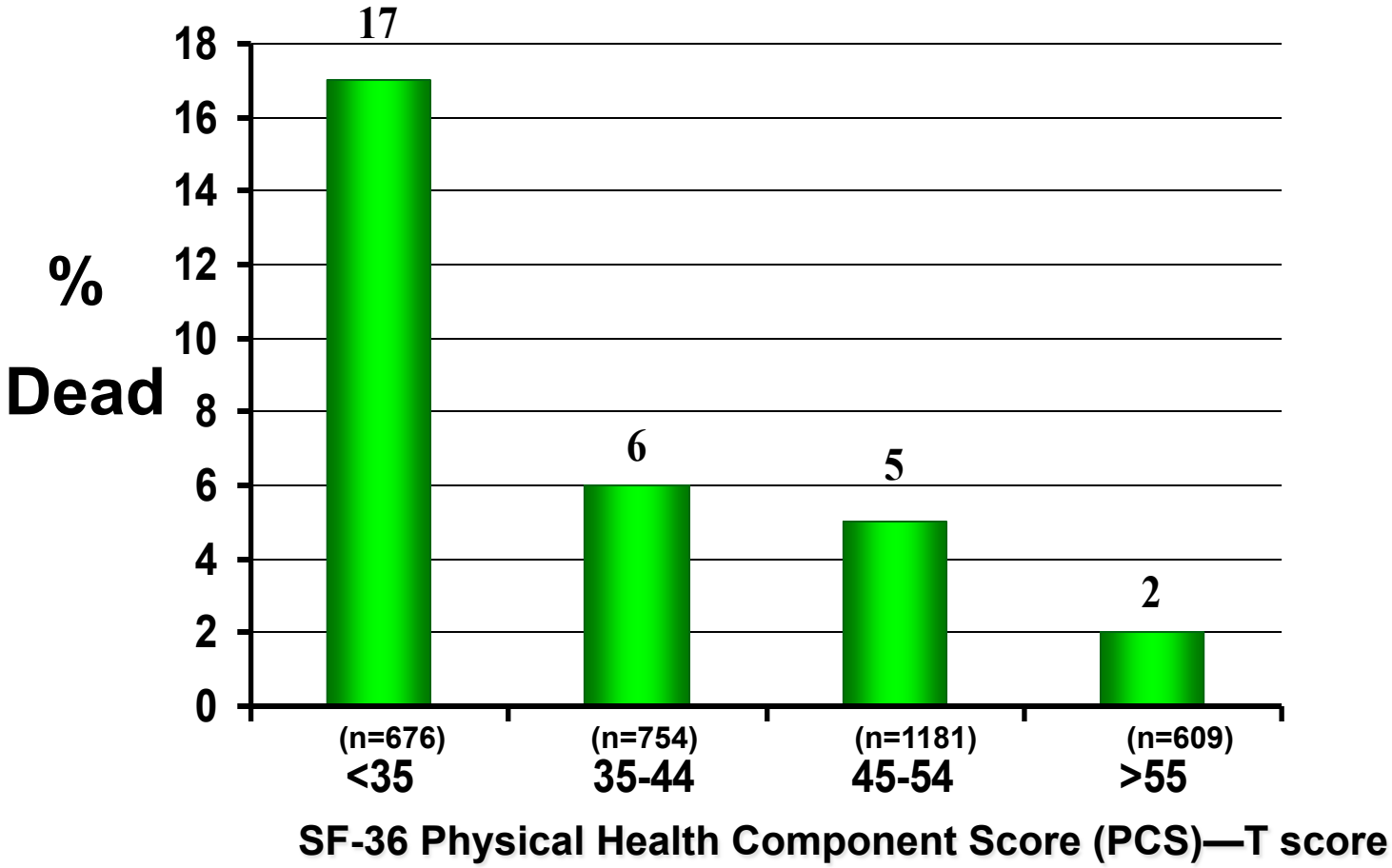
Treatment Impact on Physical Health



Treatment Impact on Mental Health



PCS Predictive of 5-Year Mortality



Ware et al. (1994). SF-36 Physical and Mental Health Summary Scales: A User's Manual.

Weights

Summary scores for SF-36 derived from uncorrelated (orthogonal) two factor (physical and mental health) solution

PCS-z = (PF-z*.42) + (RP-z*.35) + (BP-z*.32) + (GH-z*.25) + (EN-z*.03) + (SF-z*-.01) + (RE-z*-.19) + (MH-z*-.22)

MCS-z = (PF-z*-.23) + (RP-z*-.12) + (BP-z*-.10) + (GH-z*-.12) + (EN-z*.24) + (SF-z*.27) + (RE-z*.43) + (MH-z*.48)

Debate About Summary Scores

•Taft, C., Karlsson, J., & Sullivan, M. (2001). Do SF-36 component score accurately summarize subscale scores? Quality of Life Research, 10, 395-404.

•Ware, J. E., & Kosinski, M. (2001). Interpreting SF-36 summary health measures: A response. Quality of Life Research, 10, 405-413.

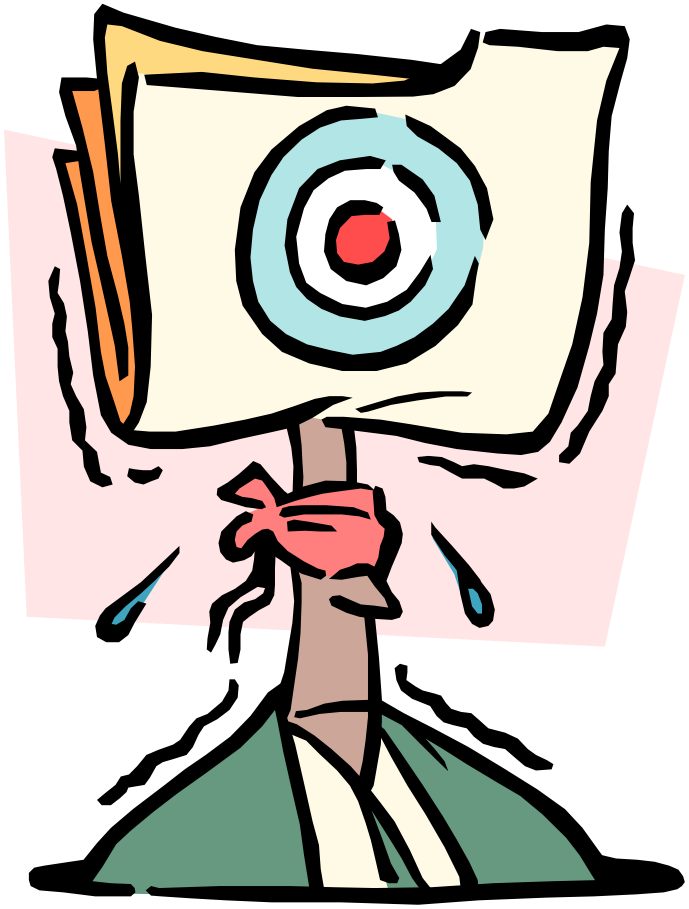
•Taft, C., Karlsson, J., & Sullivan, M. (2001). Reply to Drs Ware and Kosinski. Quality of Life Research, 10, 415-420.

536 Primary Care Patients Initiating Antidepressant Tx

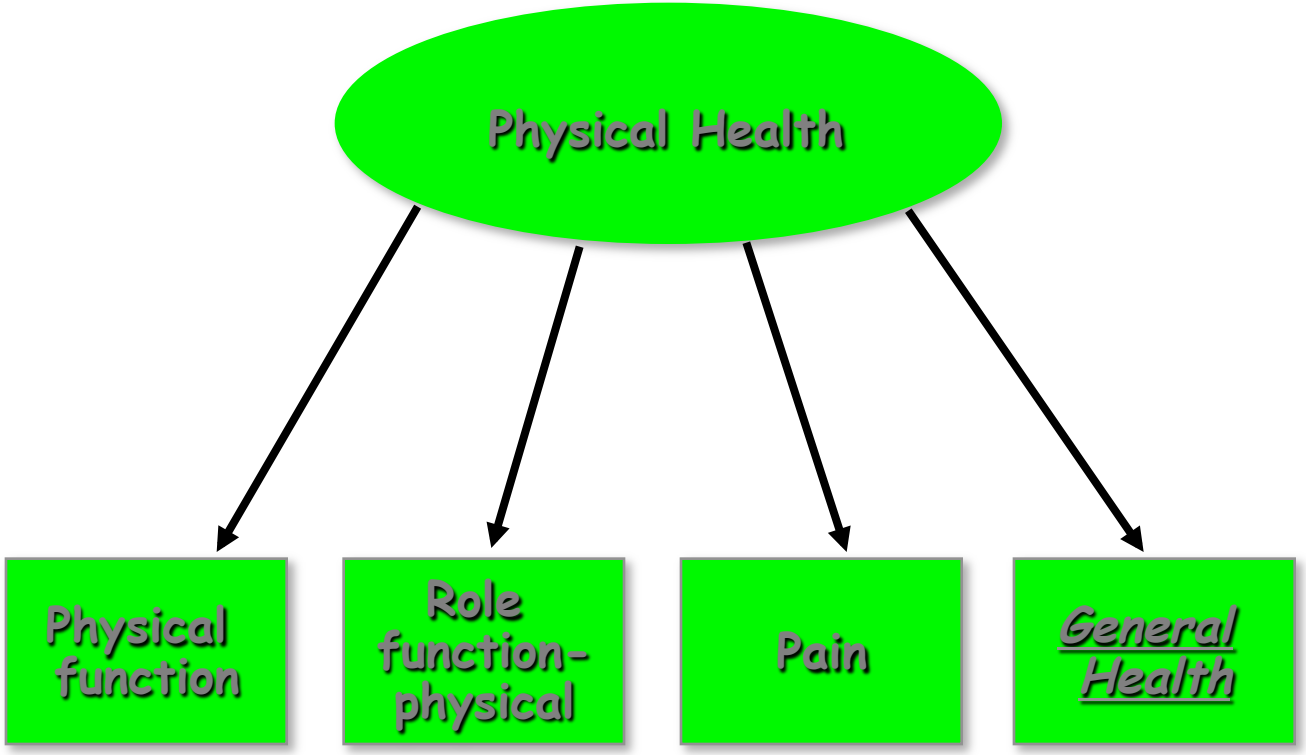
- 3-month improvements in physical functioning, role—physical, pain, and general health perceptions ranging from 0.28 to 0.49 SDs.

- Yet SF-36 PCS did not improve.

- *Simon et al. (Med Care, 1998)*



Four scales improve 0.28-0.49 SD, but physical health summary score doesn't change



n = 194 with Multiple Sclerosis

- Lower scores than general population on
  - Emotional well-being (↓ 0.3 SD)
  - Role—emotional (↓ 0.7 SD)
  - Energy (↓1.0 SD)
  - Social functioning (↓1.0 SD)
- Yet SF-36 MCS was only 0.2 SD lower.

- *Nortvedt et al. (Med Care, 2000)*

Four scales 0.3-1.0 SD lower, but MCS only 0.2 SD lower

Farivar et al. alternative weights

PCS_z = (PF_z * .20) + (RP_z * .31) +
(BP_z * .23) + (GH_z * .20) +
(EF_z * .13) + (SF_z * .11) +
(RE_z * .03) + (EW_z * -.03)

MCS_z = (PF_z * -.02) + (RP_z * .03) +
(BP_z * .04) + (GH_z * .10) +
(EF_z * .29) + (SF_z * .14) +
(RE_z * .20) + (EW_z * .35)

Targeted HRQOL Measures

- Designed to be relevant to particular group.
- Sensitive to small, but clinically-important changes.
- More familiar and actionable for clinicians.
- Enhance respondent cooperation.

Kidney-Disease Targeted Items

- During the last 30 days, to what extent were you bothered by each of the following.

  - Cramps during dialysis ?
  - Feeling washed out or drained?

  Not at all bothered
  Somewhat bothered
  Moderately bothered
  Very much bothered
  Extremely bothered

IBS-Targeted Item

- During the last 4 weeks, how often were you angry about your irritable bowel syndrome?

None of the time

A little of the time

Some of the time

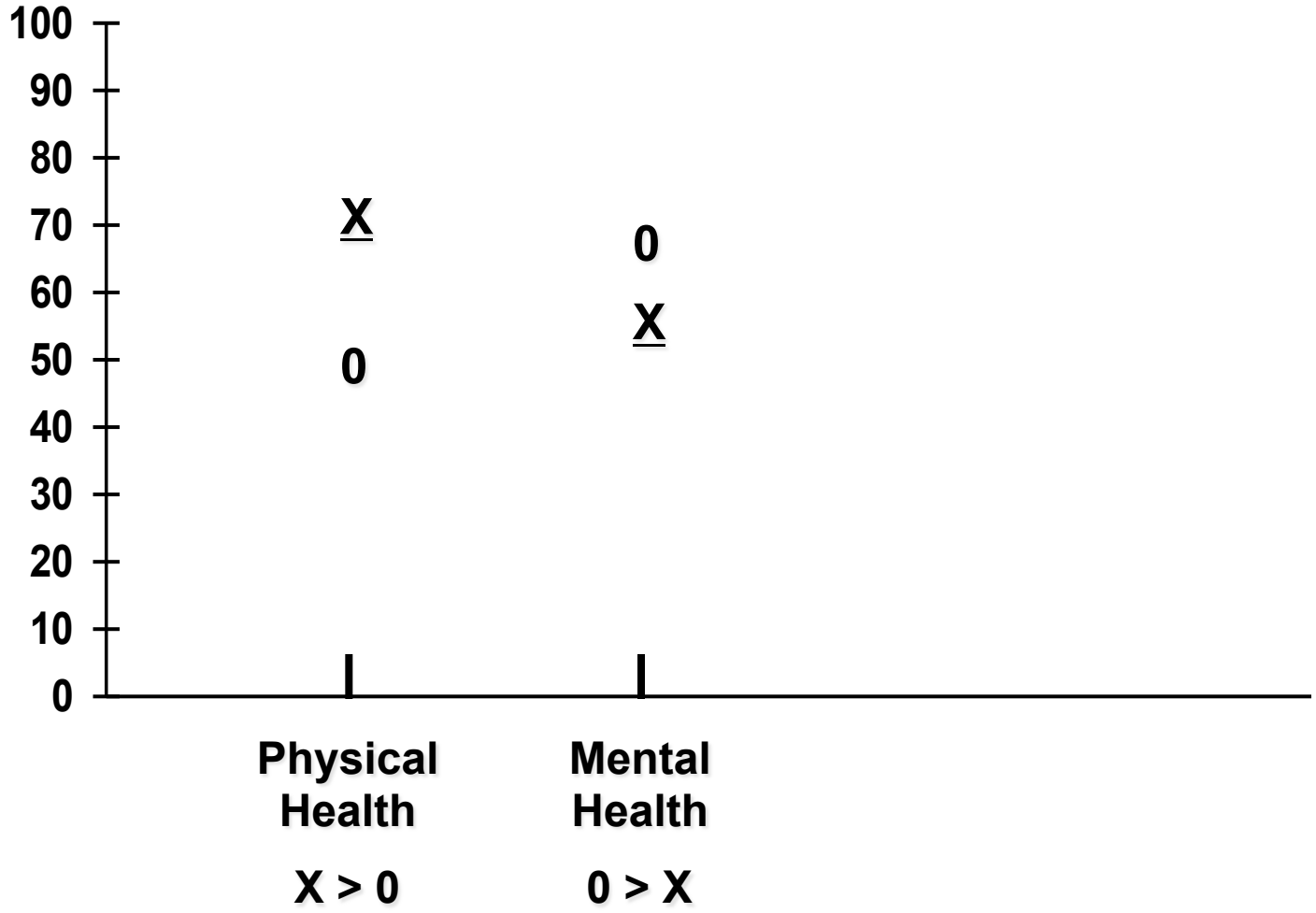
Most of the time

All of the time

Cost-Effective Health Care

$$\frac{\text{Cost } \boxed{?}}{\text{Effectiveness } \boxed{?}}$$

Is New Treatment (X) Better Than Standard Care (O)?



In general, how would you rate your health?

Poor
Fair
Good
Very Good
Excellent

Is Medicine Related to Worse HRQOL?

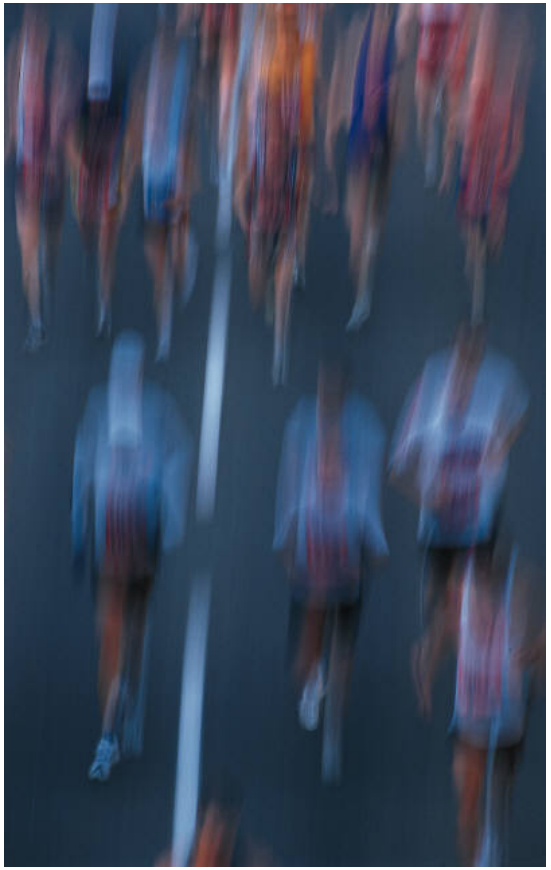
Person	Medication Use	HRQOL (0-100 scale)
1	No	dead
2	No	dead
3	No	50
4	No	75
5	No	100
6	Yes	0
7	Yes	25
8	Yes	50
9	Yes	75
10	Yes	100

Group	n	HRQOL
No Medicine	3	75
Yes Medicine	5	50

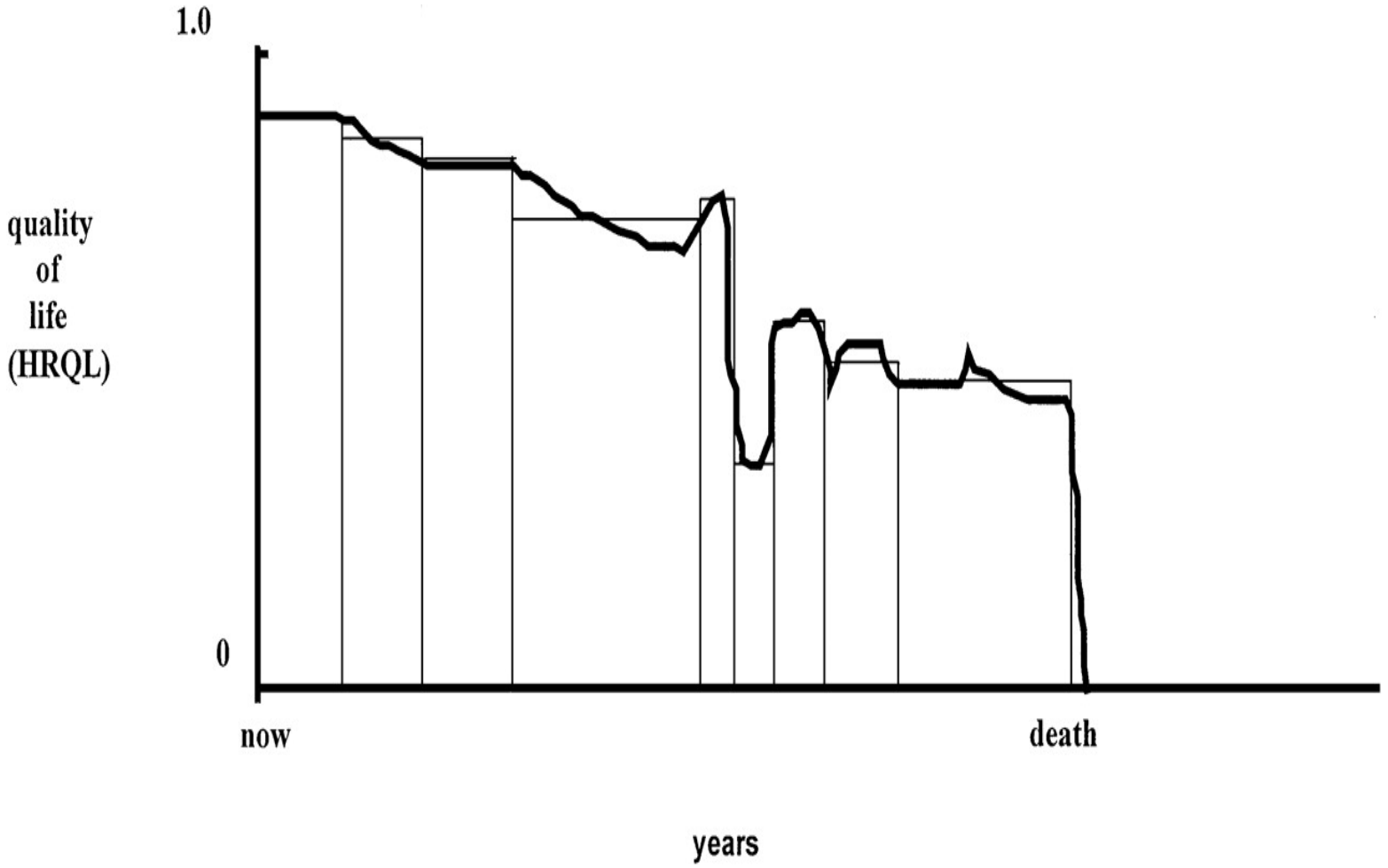
Survival Analysis

Marathoner 1.0

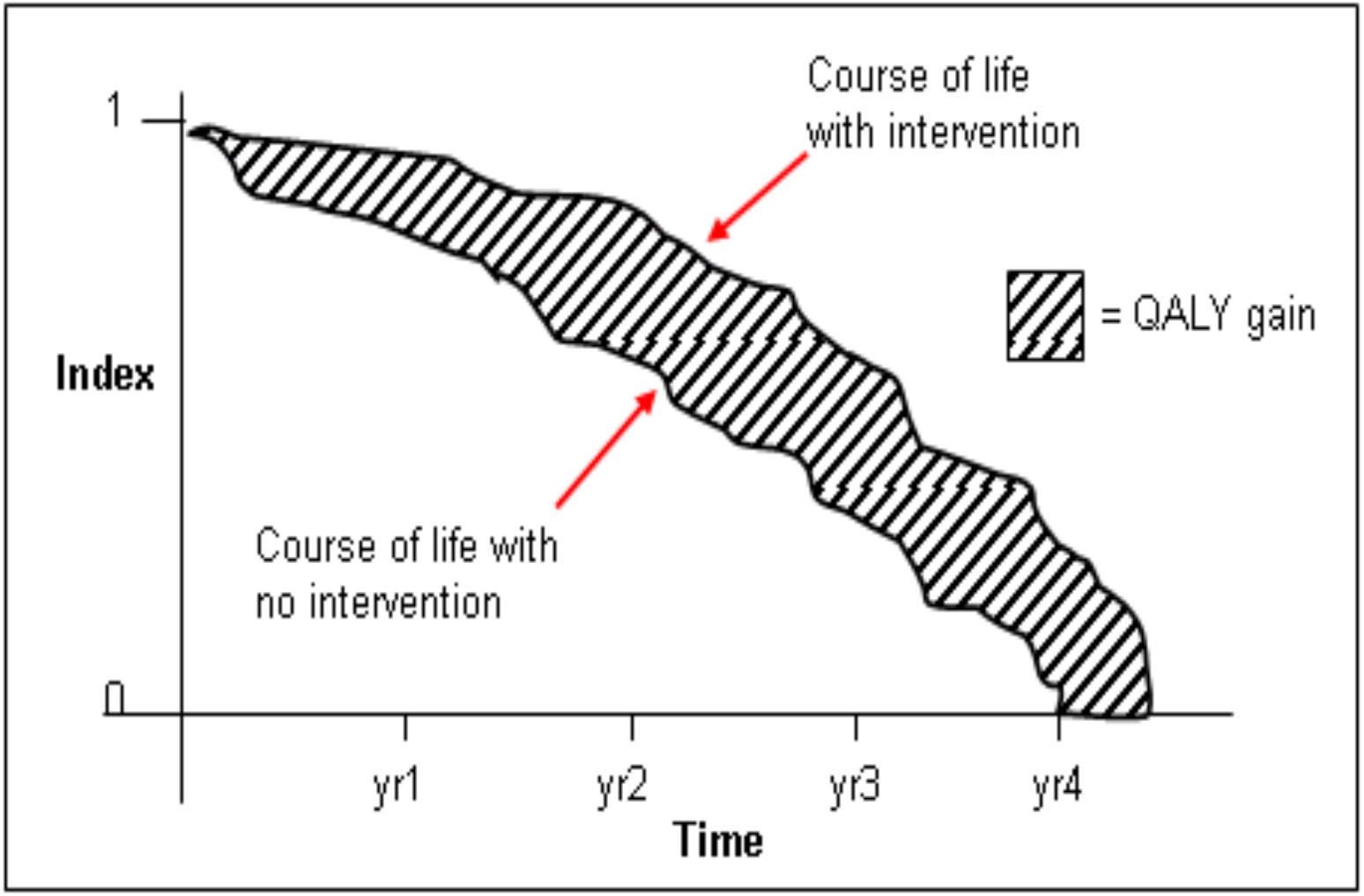
Person in coma 1.0



Quality of Life for Individual Over Time



http://www.ukmi.nhs.uk/Research/pharma_res.asp



Your own health state today

By placing a tick in one box in each group below, please indicate which statement best describes your own health state today.

Do not tick more than one box in each group.

Mobility
- I have no problems in walking about ☐
- I have some problems in walking about ☐
- I am confined to bed ☐

Self-Care
- I have no problems with self-care ☐
- I have some problems washing and dressing myself ☐
- I am unable to wash or dress myself ☐

Usual Activities (eg. work, study, housework, family or leisure activities)
- I have no problems with performing my usual activities ☐
- I have some problems with performing my usual activities ☐
- I am unable to perform my usual activities ☐

Pain/Discomfort
- I have no pain or discomfort ☐
- I have moderate pain or discomfort ☐
- I have extreme pain or discomfort ☐

Anxiety/Depression
- I am not anxious or depressed ☐
- I am moderately anxious or depressed ☐
- I am extremely anxious or depressed ☐

On each dimension, respondent gets three choices of level.

Health state 424421 (0.59)

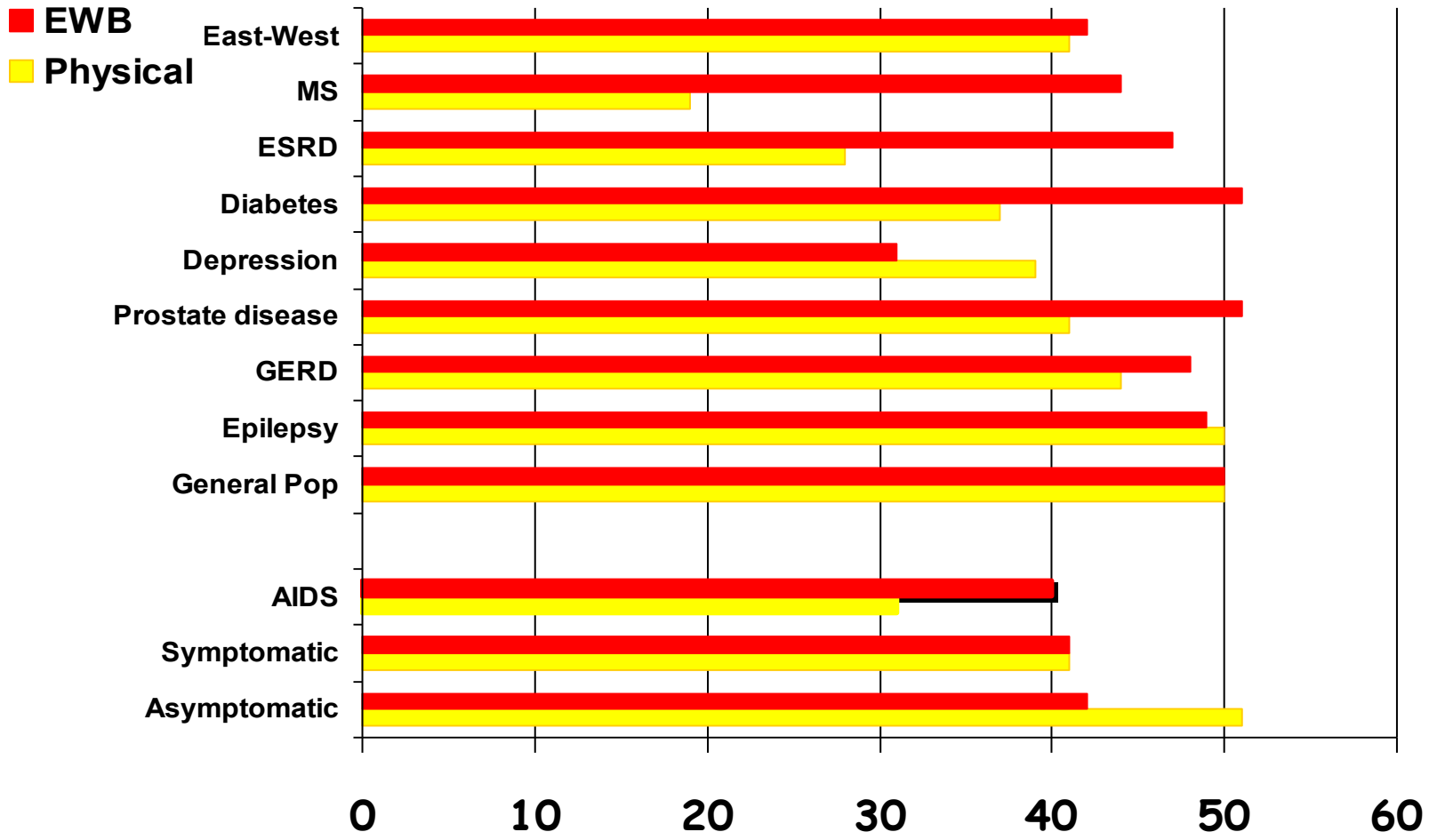
- Your health limits you a lot in moderate activities (such as moving a table, pushing a vacuum cleaner, bowling or playing golf)
- You are limited in the kind of work or other activities as a result of your physical health
- Your health limits your social activities (like visiting friends, relatives etc.) most of the time.
- You have pain that interferes with your normal work (both outside the home and housework) moderately
- You feel tense or downhearted and low a little of the time.
- You have a lot of energy all of the time

Part II: Evaluating Individual Change

Individual Change

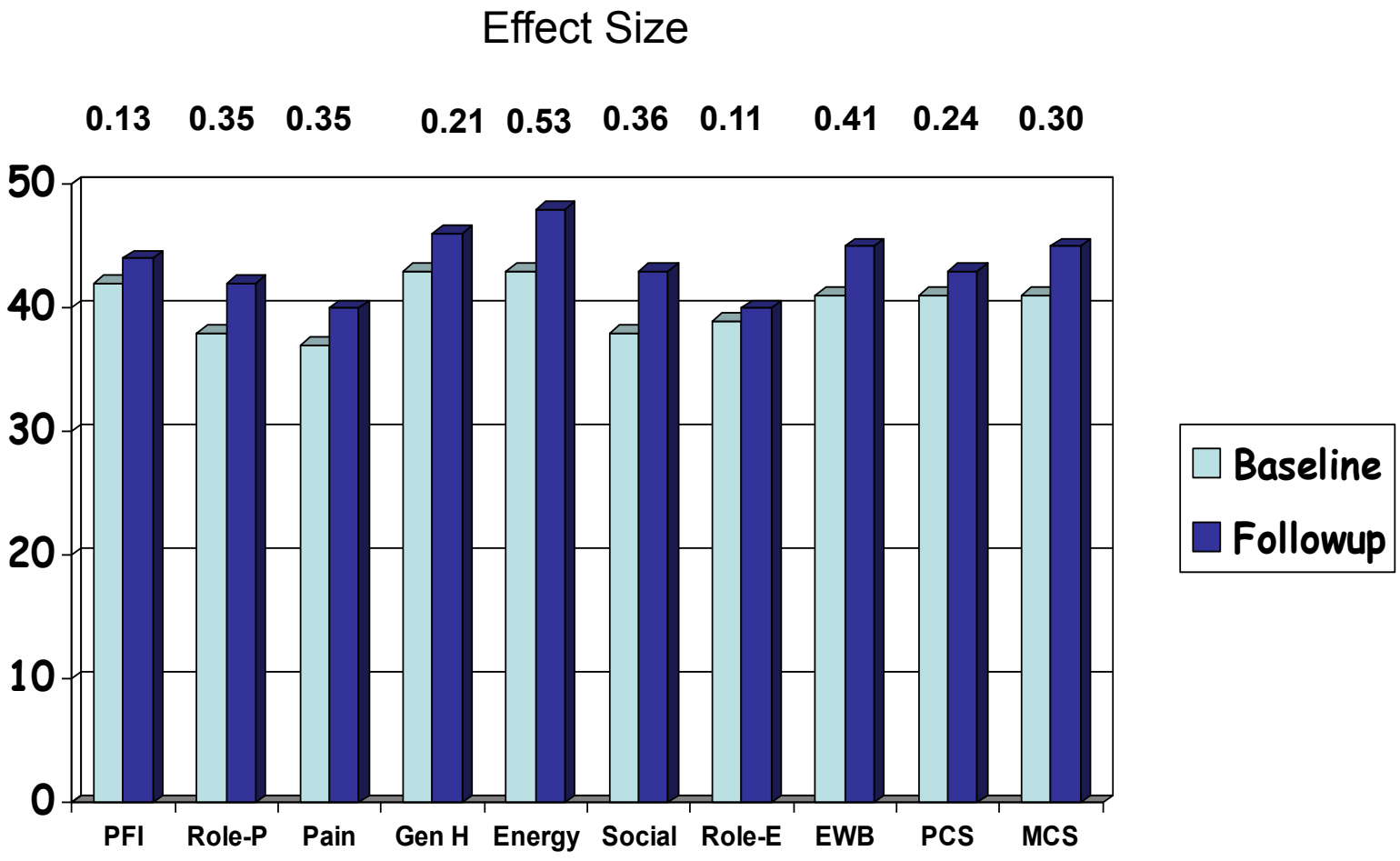
- Motivation
  - Knowing how many patients benefit from group intervention, or
  - Tracking progress on individual patients
- Sample
  - 54 patients
  - Average age = 56; 84% white; 58% female
- Method
  - Self-administered SF-36 version 2 at baseline and at end of therapy (about 6 weeks later).

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



Hays et al. (2000), American Journal of Medicine

Change in SF-36 Scores Over Time



t-test for within group change

$$X_D/(SD_d/n^{1/2})$$

X_D = is mean difference, SD_d = standard deviation of difference

Significance of Group Change (T-scores)

	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

Reliable Change Index

$(X_2 – X_1)/ (SEM * SQRT [2])$

$SEM = SD_b * (1- reliability)^{1/2}$

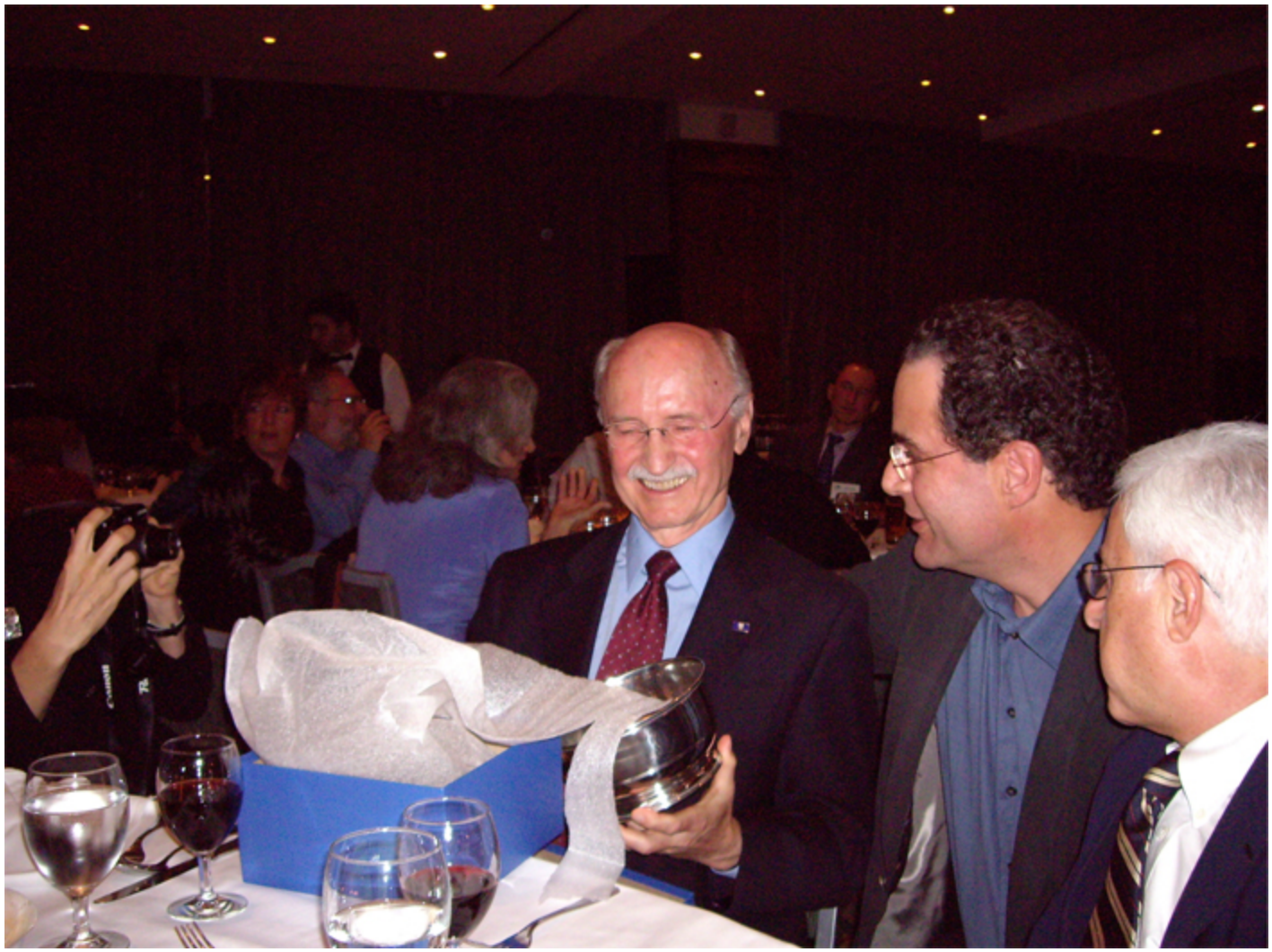
Amount of Change in Observed Score Needed for Significant Individual Change

	RCI	Effect size
PF-10	8.4	0.67
RP-4	8.4	0.72
BP-2	10.4	1.01
GH-5	13.0	1.13
EN-4	12.8	1.33
SF-2	13.8	1.07
RE-3	9.7	0.71
EWB-5	13.4	1.26
PCS	7.1	0.62
MCS	9.7	0.73

Significant Change for 54 Cases

	% 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%

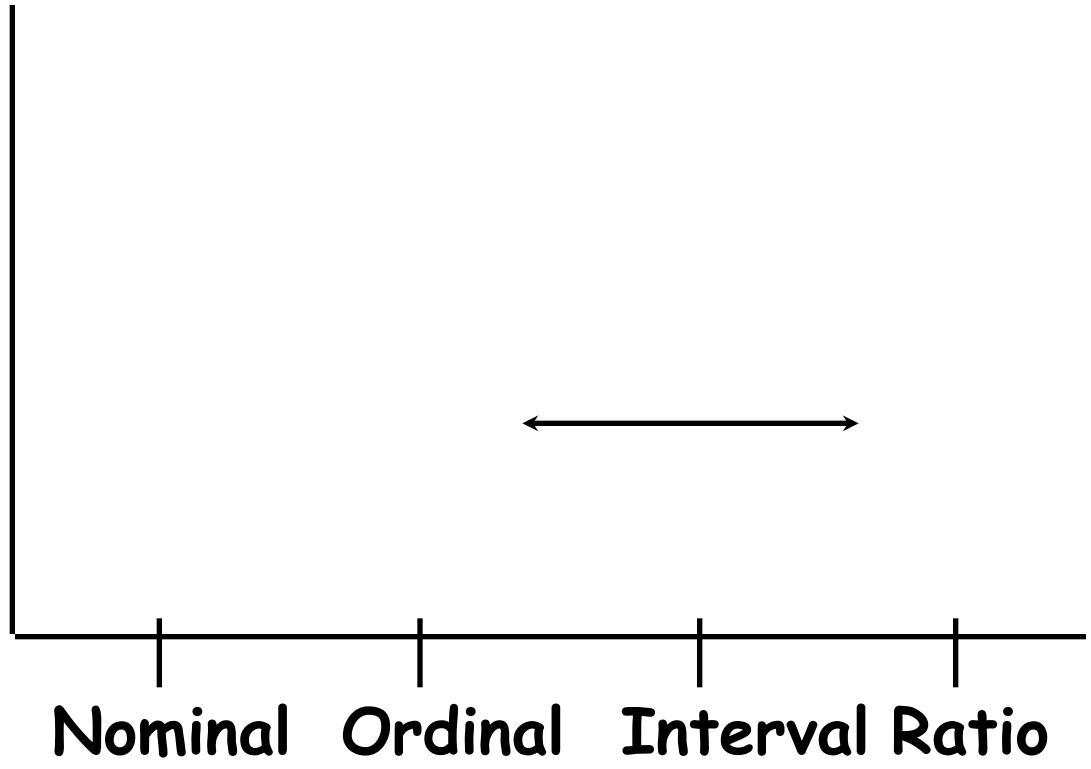
Part III: Psychometric Properties of a Good Measure



Scales of Measurement and Their Properties

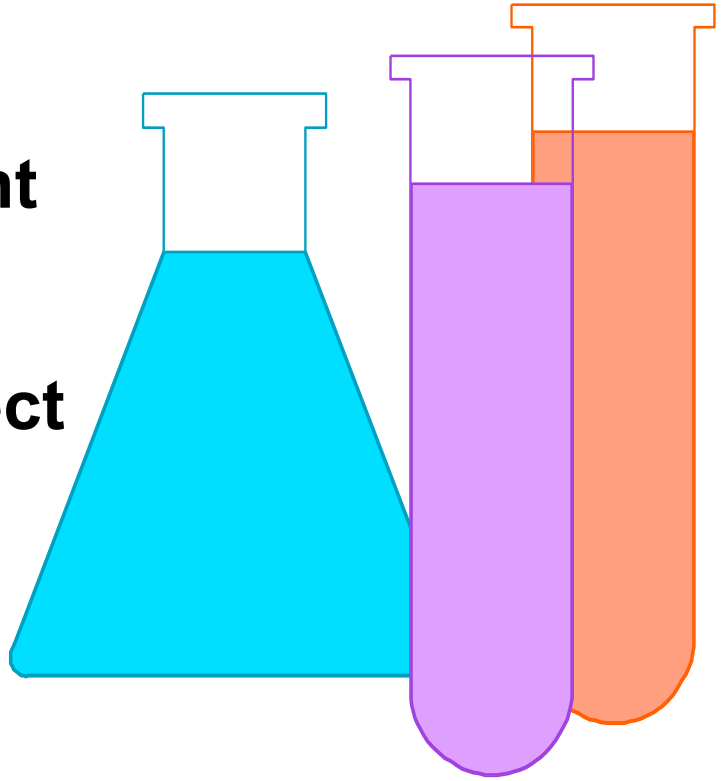
Type of Scale	Property of Numbers: Rank Order	Equal Interval	Absolute 0
Nominal			
Ordinal	+		
Interval	+	+	
Ratio	+	+	+

Measurement Range for Health Outcome Measures



What's a Good Measure?

- **Same person gets same score (reliability)**
- **Different people get different scores (validity)**
- **People get scores you expect (validity)**
- **Practical to use (feasibility)**



Indicators of Acceptability

- Unit non-response
- Item non-response

How many items can people complete in 15 minutes?

Variability

- All scale levels are represented
- Distribution approximates bell-shaped "normal"

Measurement Error

observed = true score + systematic error + random error

(bias)

Measurement Error is One Source of Data Collection Error

- **Coverage Error**
  Does each person in population have an equal chance of selection?
- **Sampling Error**
  Are only some members of the population sampled?
- **Nonresponse Error**
  Do people in the sample who respond differ from those who do not?

Flavors of Reliability

- Test-retest (administrations)
- Intra-rater (raters)
- Internal consistency (items)

Intraclass Correlation and Reliability

Model	Reliability	Intraclass Correlation
One-way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$
Two-way fixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} + MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$

BMS = Between Ratee Mean Square
WMS = Within Mean Square
JMS = Item or Rater Mean Square
EMS = Ratee x Item (Rater) Mean Square

Spearman-Brown Prophecy Formula

$$alpha_y = \left(\frac{N \bullet alpha_x}{1 + (N - 1) * alpha_x}\right)$$

N = how much longer scale y is than scale x

Example Spearman-Brown Calculations

MHI-18

$$\frac{18/32\ (0.98)}{(1+(18/32\ -1)*0.98}$$

= 0.55125/0.57125 = 0.96

Reliability Minimum Standards

- 0.70 or above (for group comparisons)
- 0.90 or higher (for individual assessment)
  - SEM = SD (1- reliability)$^{1/2}$

Hypothetical Multitrait/ Multi-Item Correlation Matrix

	Trait #1	Trait #2	Trait #3
Item #1	0.80*	0.20	0.20
Item #2	0.80*	0.20	0.20
Item #3	0.80*	0.20	0.20
Item #4	0.20	0.80*	0.20
Item #5	0.20	0.80*	0.20
Item #6	0.20	0.80*	0.20
Item #7	0.20	0.20	0.80*
Item #8	0.20	0.20	0.80*
Item #9	0.20	0.20	0.80*

*Item-scale correlation, corrected for overlap.

Construct Validity

- Does measure relate to other measures in ways consistent with hypotheses?

- Responsiveness to change including minimally important difference

MTMM.EXE (2.3): Multitrait-Multimethod Program

Hayashi, T., & Hays, R. D. (1987). A microcomputer program for analyzing multitrait-multimethod matrices. Behavior Research Methods, Instruments, & Computers, 19 (3), 345-348.

Correlation Matrix Input Is As Follows:

Kobayashi PEDSql 2007

N = 790; DFS = 787

	METHOD	1				2			
	TRAIT	1	2	3	4	1	2	3	4
1.	1.PHYSICAL	1.00							
	2.EMOTIONA	.48	1.00						
	3.SOCIAL F	.43	.52	1.00					
	4.SCHOOL F	.46	.42	.39	1.00				
2.	1.PHYSICAL	[.19]	.13	.13	.17	1.00			
	2.EMOTIONA	.27	[.32]	.20	.24	.44	1.00		
	3.SOCIAL F	.22	.26	[.34]	.21	.45	.57	1.00	
	4.SCHOOL F	.18	.21	.22	[.41]	.39	.52	.57	1.00

(Total Z = 1.31 Mean Z = .33)

Average convergent validity correlation is .317
Average off-diagonal correlation is .345

==========

Responsiveness to Change and Minimally Important Difference (MID)

- HRQOL measures should be responsive to interventions that change HRQOL

- Need external indicators of change (Anchors)

  - mean change in HRQOL scores among people who have changed (“minimal” change for MID).

Self-Report Indicator of Change

- Overall has there been any change in your asthma since the beginning of the study?

  Much improved; Moderately improved; Minimally improved

  No change

  Much worse; Moderately worse; Minimally worse

Clinical Indicator of Change

- "changed" group = seizure free (100% reduction in seizure frequency)

- "unchanged" group = <50% change in seizure frequency

Effect Size

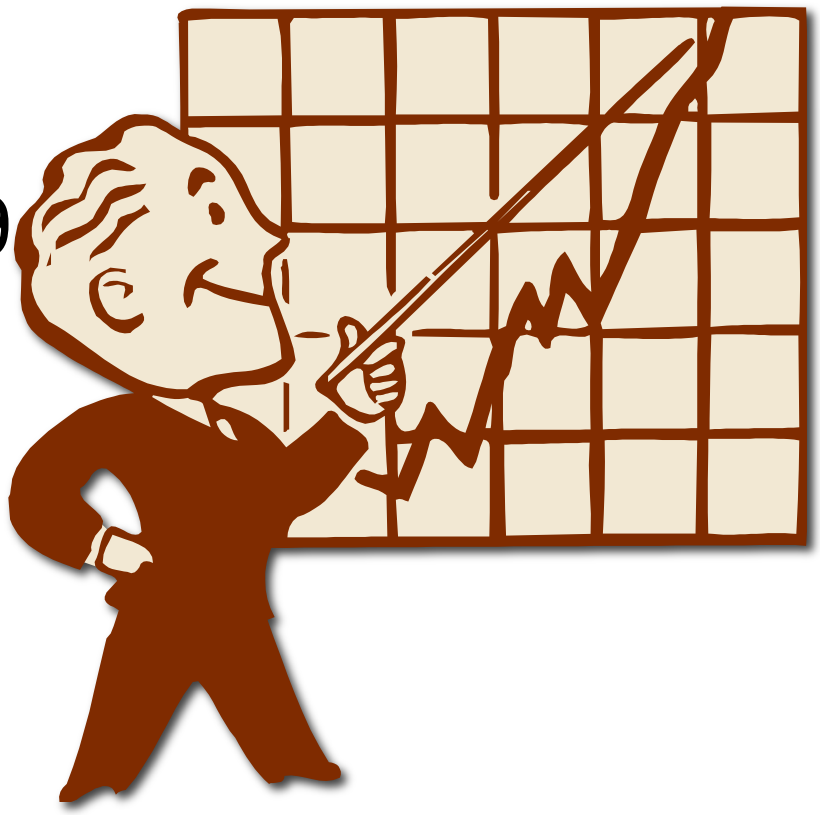
Effect size (ES) = D/SD

D = raw score change in “changed” group;

SD = baseline SD;

Effect Size Benchmarks

- Small: 0.20->0.49
- Moderate: 0.50->0.79
- Large: 0.80 or above



Responsiveness Indices

(1) Effect size (ES) = D/SD

(2) Standardized Response Mean (SRM) = D/SD†

(3) Guyatt responsiveness statistic (RS) = D/SD‡

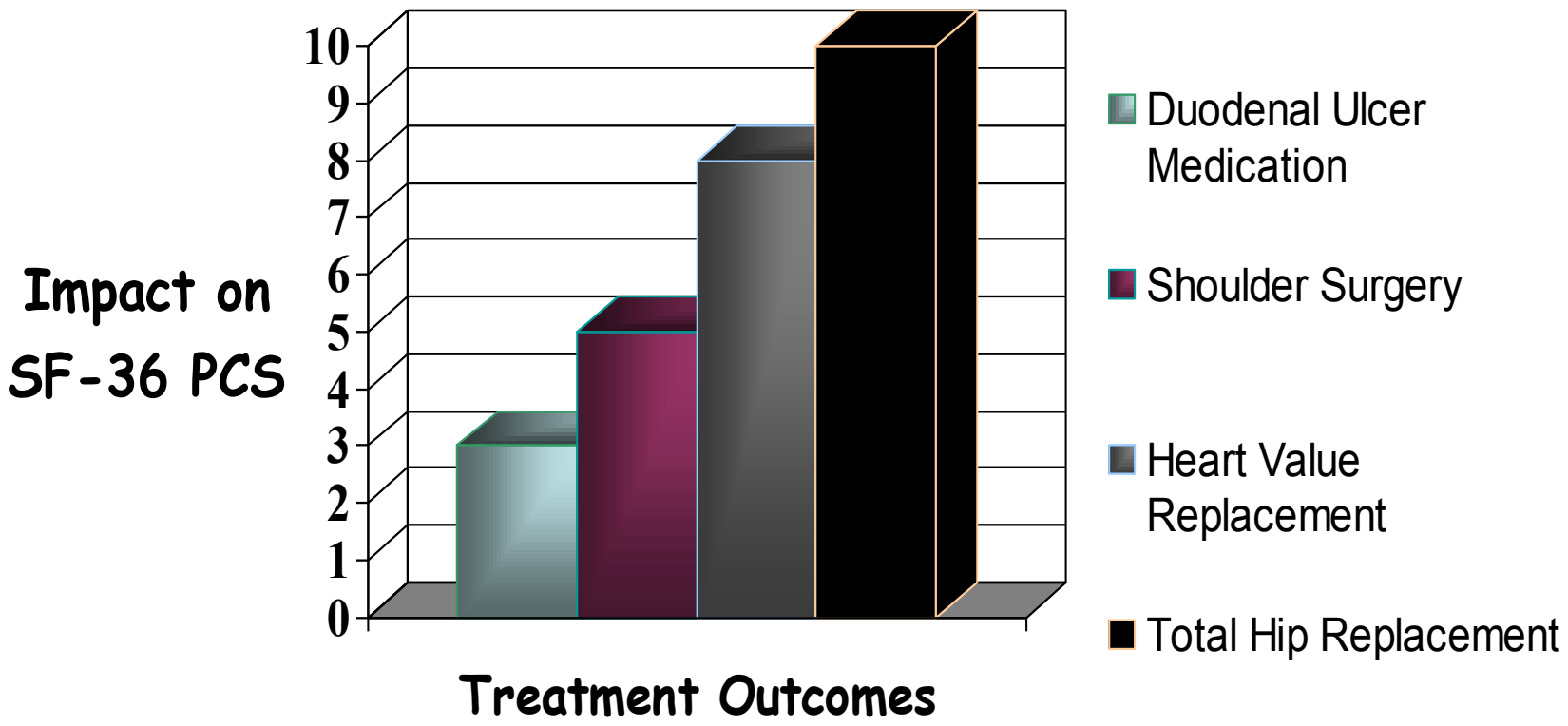
D = raw score change in "changed" group;

SD = baseline SD;

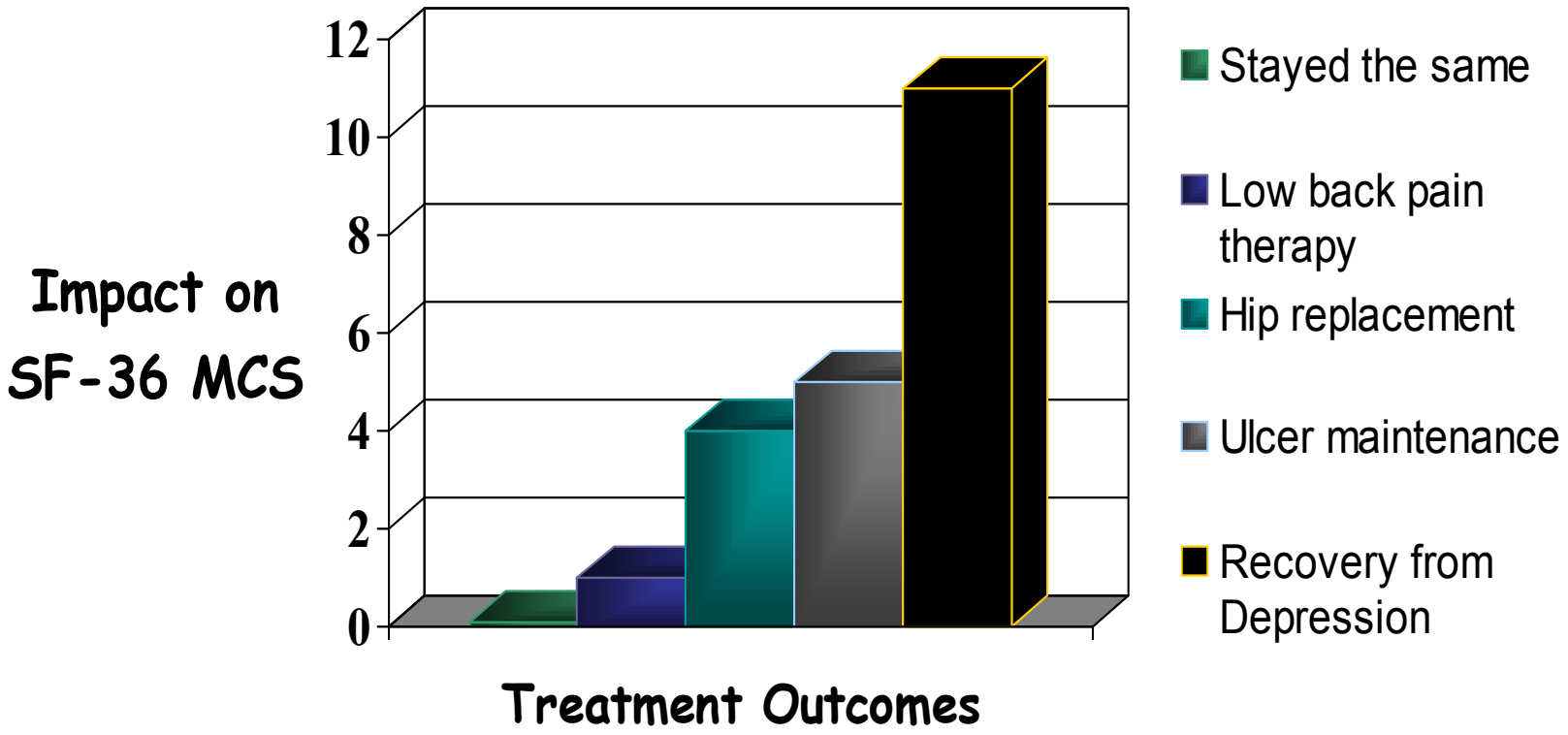
SD† = SD of D;

SD‡ = SD of D among "unchanged"

Treatment Impact on PCS



Treatment Impact on MCS



Minimally Important Difference (MID)

- One can observe a difference between two groups or within one group over time that is statistically significance but small.
- With a large enough sample size, even a tiny difference could be statistically significant.
- The MID is the smallest difference that we care about.

"Distribution-Based Estimate" of MID is not an estimate

- Distribution-based formulas
  - Effect size (ES) = D/SD
  - Standardized Response Mean (SRM) = D/SD†
  - Responsiveness statistic (RS) = D/SD‡

  SD = baseline SD; SD† = SD of D; SD‡ = SD of D among "unchanged"

- Raw score difference derived from prior information about the MID
  - e.g., $D_{measure} = ES * SD_{measure}$

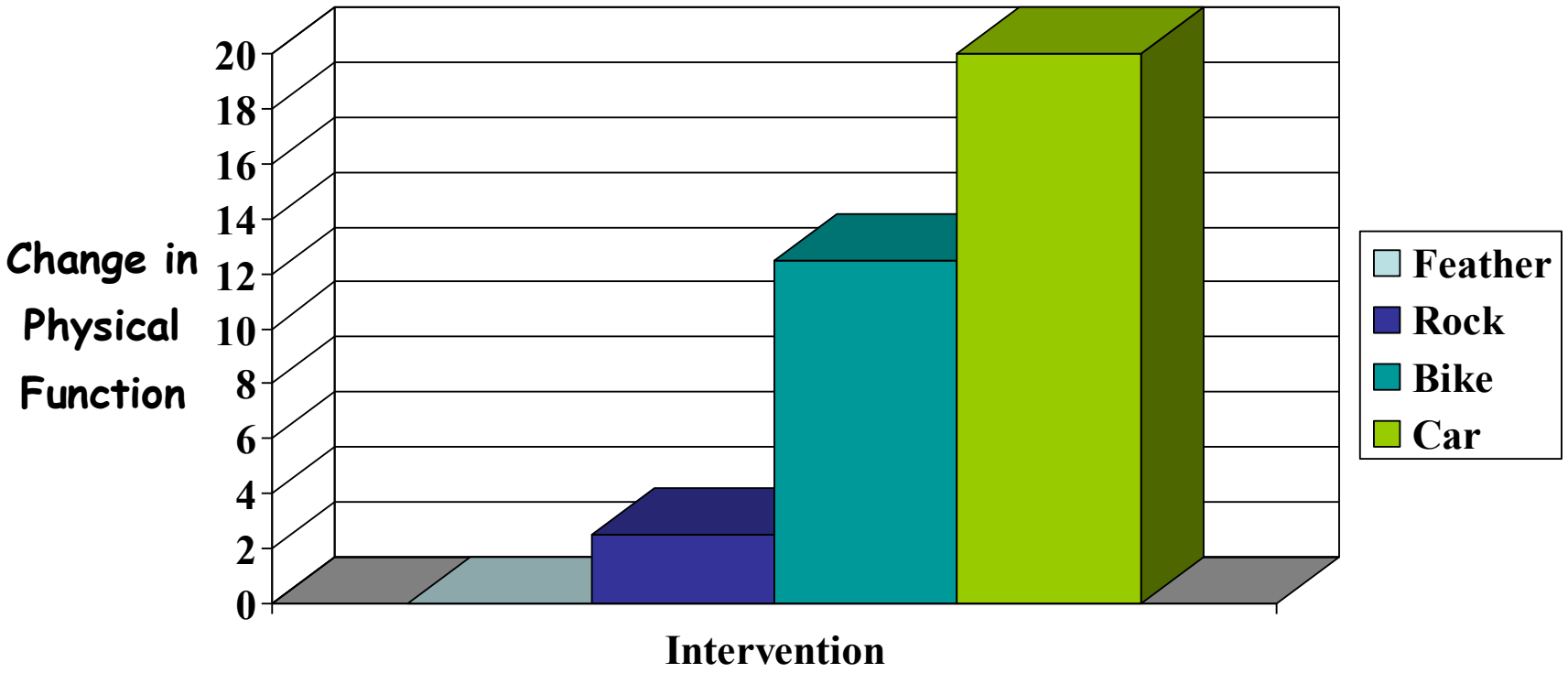
Standard Error of Measurement

- SEM = SD * SQRT (1-reliability)
- 95% CI = Estimated true score +/- 1.96 * SEM
- 1 SEM = 0.50 SD when reliability is 0.75

Estimating the MID

- External anchors
  - Self-report
  - Provider report
  - Clinical measure
  - Intervention
- Anchor correlated with change on target measure at 0.371 or higher
- Anchor indicates “minimal” change

Hypothetical Change in Physical Function (T-score units) by magnitude of intervention



The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

1. Vigorous activities, such as running, lifting heaving objects, participating in strenuous sports
2. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
3. Lifting or carrying groceries
4. Climbing several flights of stairs
5. Climbing one flight of stairs
6. Bending, kneeling, or stooping
7. Walking more than a mile
8. Walking several blocks
9. Walking one block
10. Bathing or dressing yourself

Yes, limited a lot (0)/Yes, limited a little (50)/No, not limited at all (100)
Mean = 87; 75th percentile = 100 for U.S. males

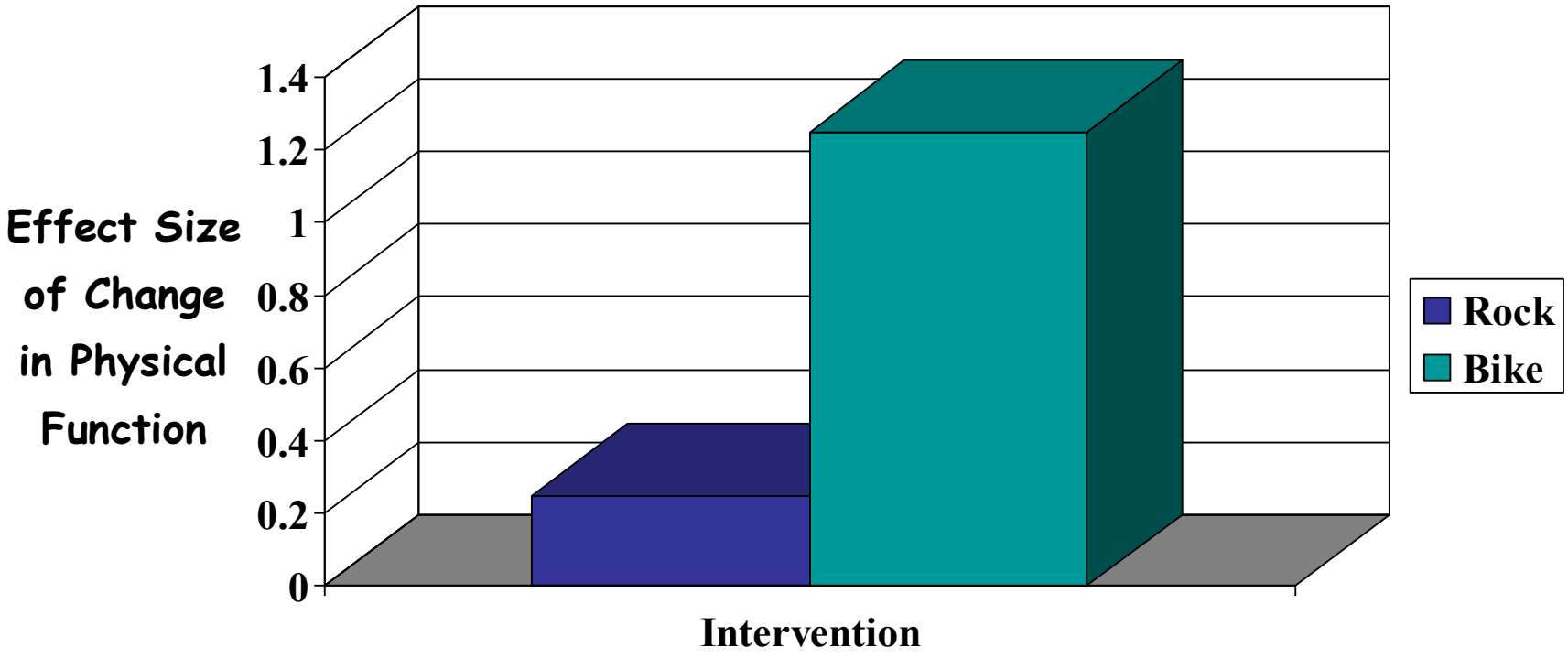
Change in Physical Function from Baseline

Baseline = 100 (U.S. males mean = 87, SD = 20)

- Hit by **Bike** causes me to be *limited a lot* in vigorous activities, *limited a little* in moderate activities, and *limited a lot* in climbing several flights of stairs. Physical functioning drops to 75 (- 1.25 SD)

- Hit by **Rock** causes me to be *limited a little* in vigorous activities and physical functioning drops to 95 (- 0.25 SD)

Getting Hit By Bike is > Minimal
Getting Hit by Rock is Closer to MID



Self-Report Anchor

- People who report a “minimal” change
- How is your physical health now compared to 4 weeks ago?

- *Much improved; Moderately Improved;*
- ***Minimally Improved****;*
- *No Change;*
- ***Minimally Worse****;*
- *Moderately Worse; Much Worse*

Example with Multiple Anchors

- 693 RA clinical trial participants evaluated at baseline and 6-weeks post-treatment.
- Five anchors:
  - 1) patient global self-report;
  - 2) physician global report;
  - 3) pain self-report;
  - 4) joint swelling;
  - 5) joint tenderness

Kosinski, M. et al. (2000). Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis and Rheumatism, 43, 1478-1487.

Patient and Physician Global Reports

- How the patient is doing, considering all the ways that RA affects him/here?

Very good (asymptomatic and no limitation of normal activities)

Good (mild symptoms and no limitation of normal activities)

Fair (moderate symptoms and limitation of normal activities)

Poor (severe symptoms and inability to carry out most normal activities)

Very poor (very severe symptoms that are intolerable and inability to carry out normal activities)

--> Improvement of 1 level over time

Global Pain, Joint Swelling and Tenderness

- 0 = no pain, 10 = severe pain; 10 centimeter visual analog scale
- Number of swollen and tender joints

-> 1-20% improvement over time

Effect Sizes (mean = 0.34) for SF-36 Changes Linked to Minimal Change in Anchors

Scale	Self-R	Clin.-R	Pain	Swell	Tender	Mean
PF	.35	.33	.34	.26	.32	.32
Role-P	.56	.52	.29	.35	.36	.42
Pain	.83	.70	.47	.69	.42	.62
GH	.20	.12	.09	.12	.04	.12
EWB	.39	.26	.25	.18	.05	.23
Role-E	.41	.28	.18	.38	.26	.30
SF	.43	.34	.28	.29	.38	.34
EF	.50	.47	.22	.22	.35	.35
PCS	.49	.48	.34	.29	.36	.39
MCS	.42	.27	.19	.27	.20	.27

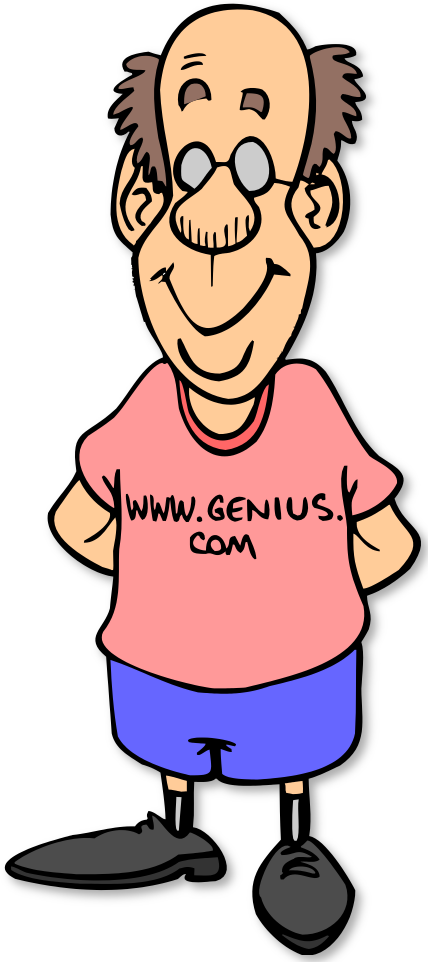
Use of “No Change” Group in Estimating MID

Group	Change #1	Change #2	Change #3
Minimal Change on Anchor	**0**	**+ 2**	**+ 4**
No Change on Anchor	Given above, not relevant	+ 2	<3
MID =	?	?	4

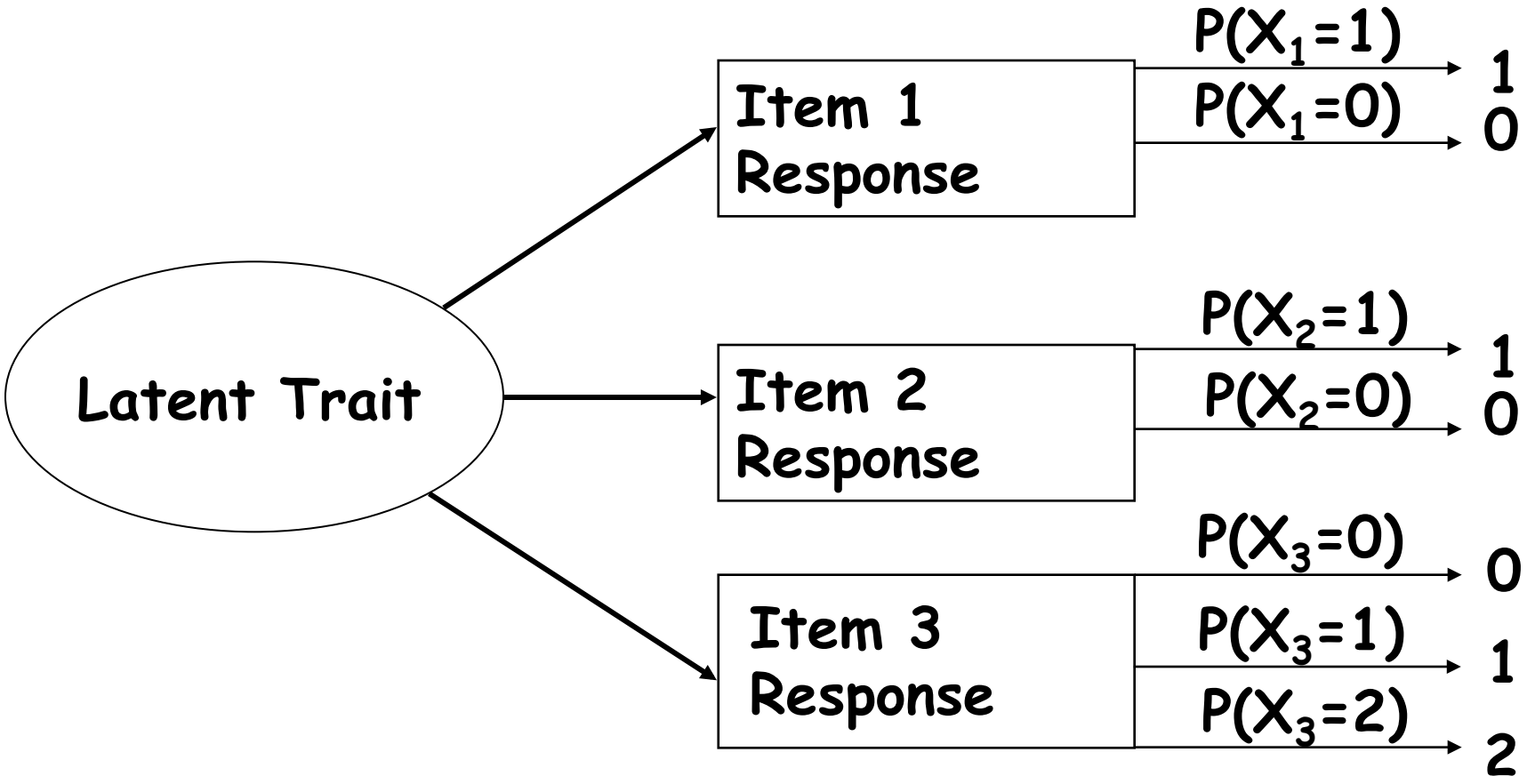
MID Summary

- Easier to conclude that a difference is clearly or obviously important than it is to say it is always unimportant.
- No single best way to estimate MID
  - Use multiple anchors
  - Use anchors that represent minimum change
- Wide variation in estimates of MID
  - Report range, inter-quartile range, and confidence intervals around mean estimates.

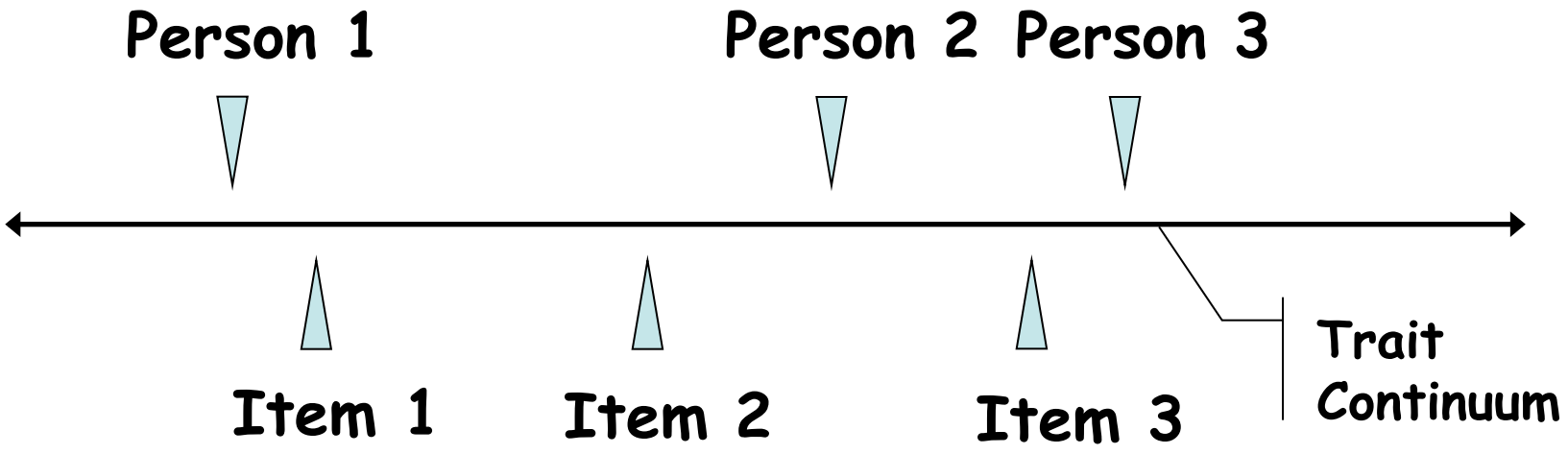
Differential Item Functioning (DIF)



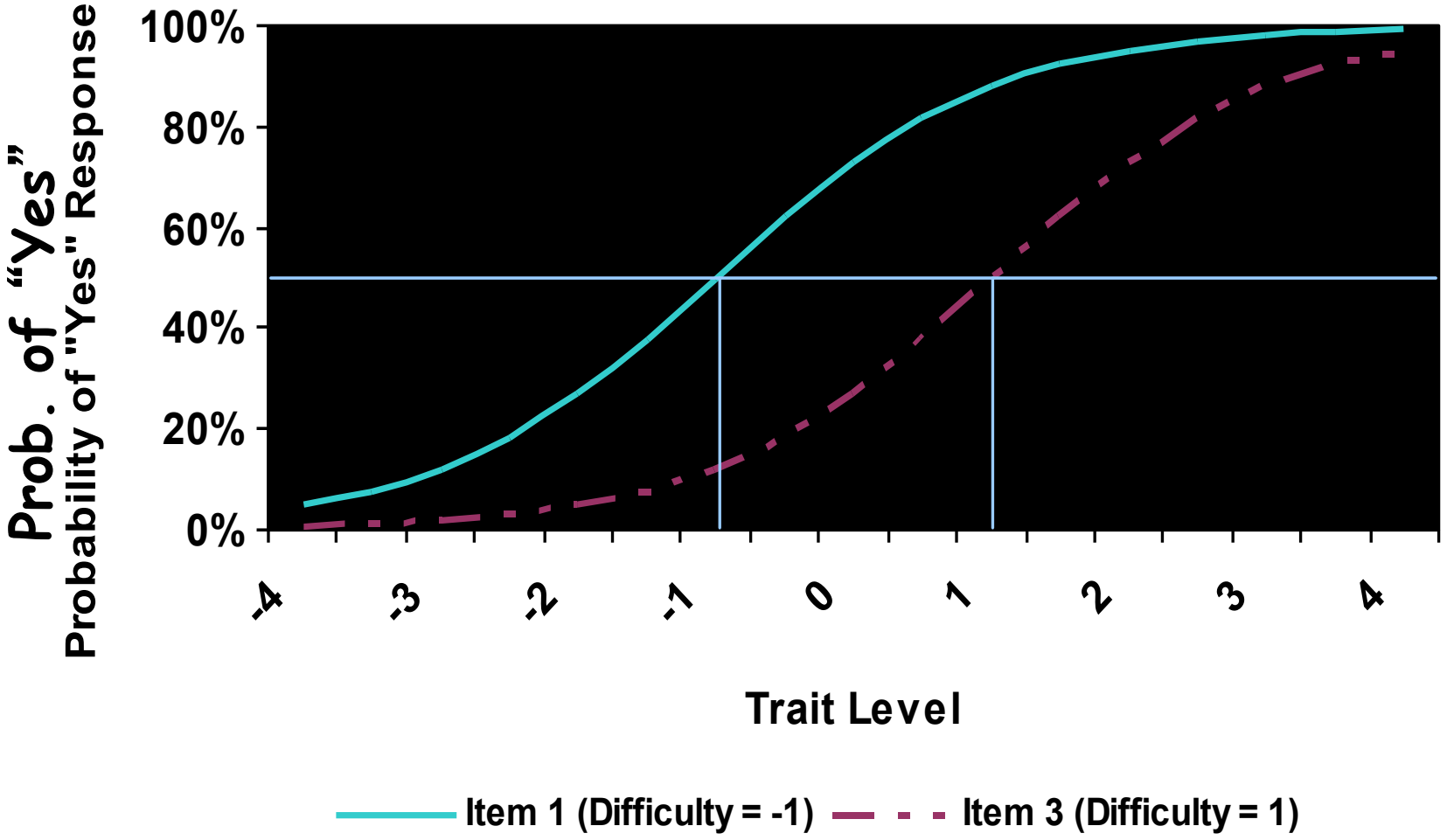
Latent Trait and Item Responses



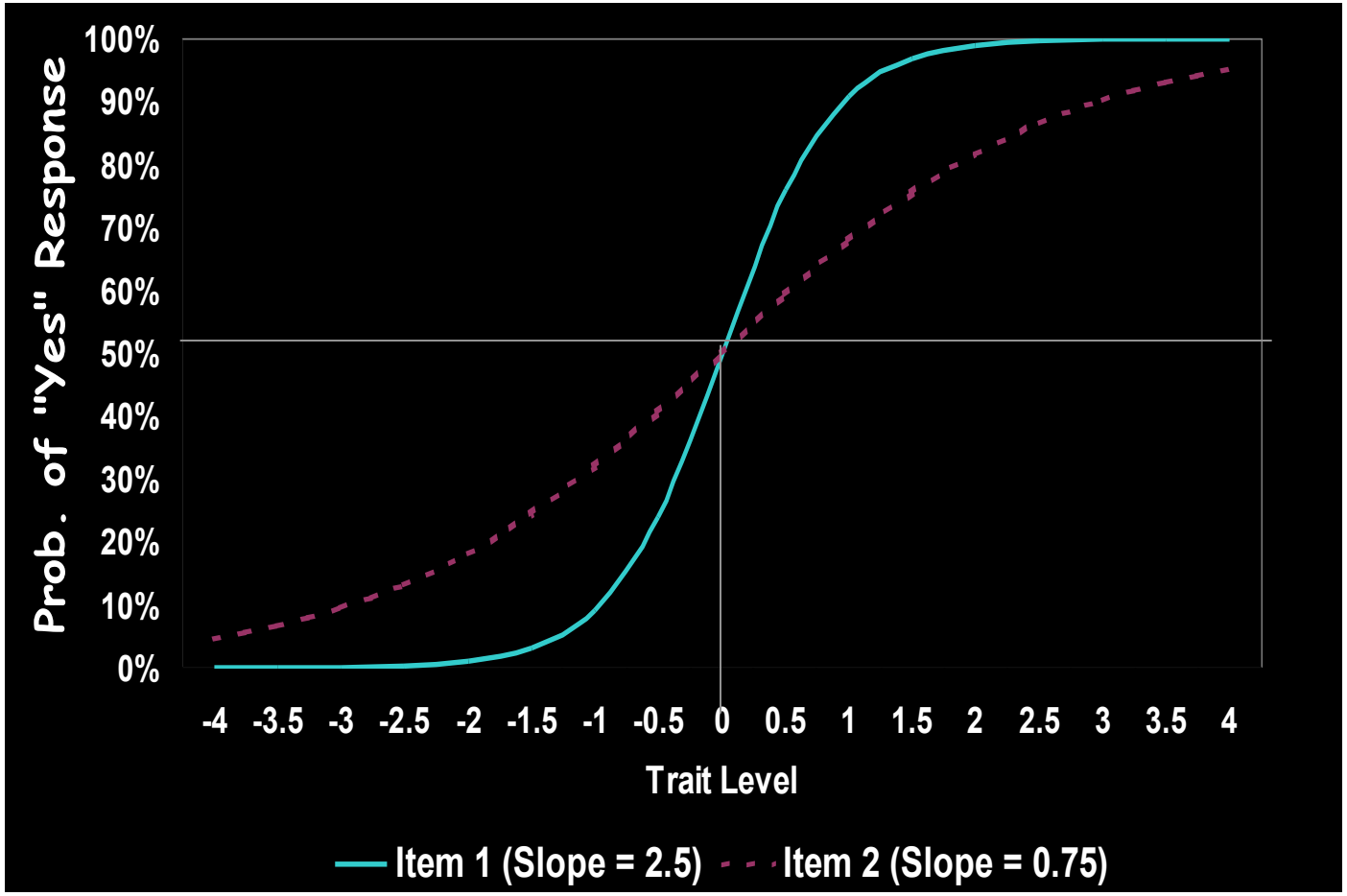
Item Responses and Trait Levels



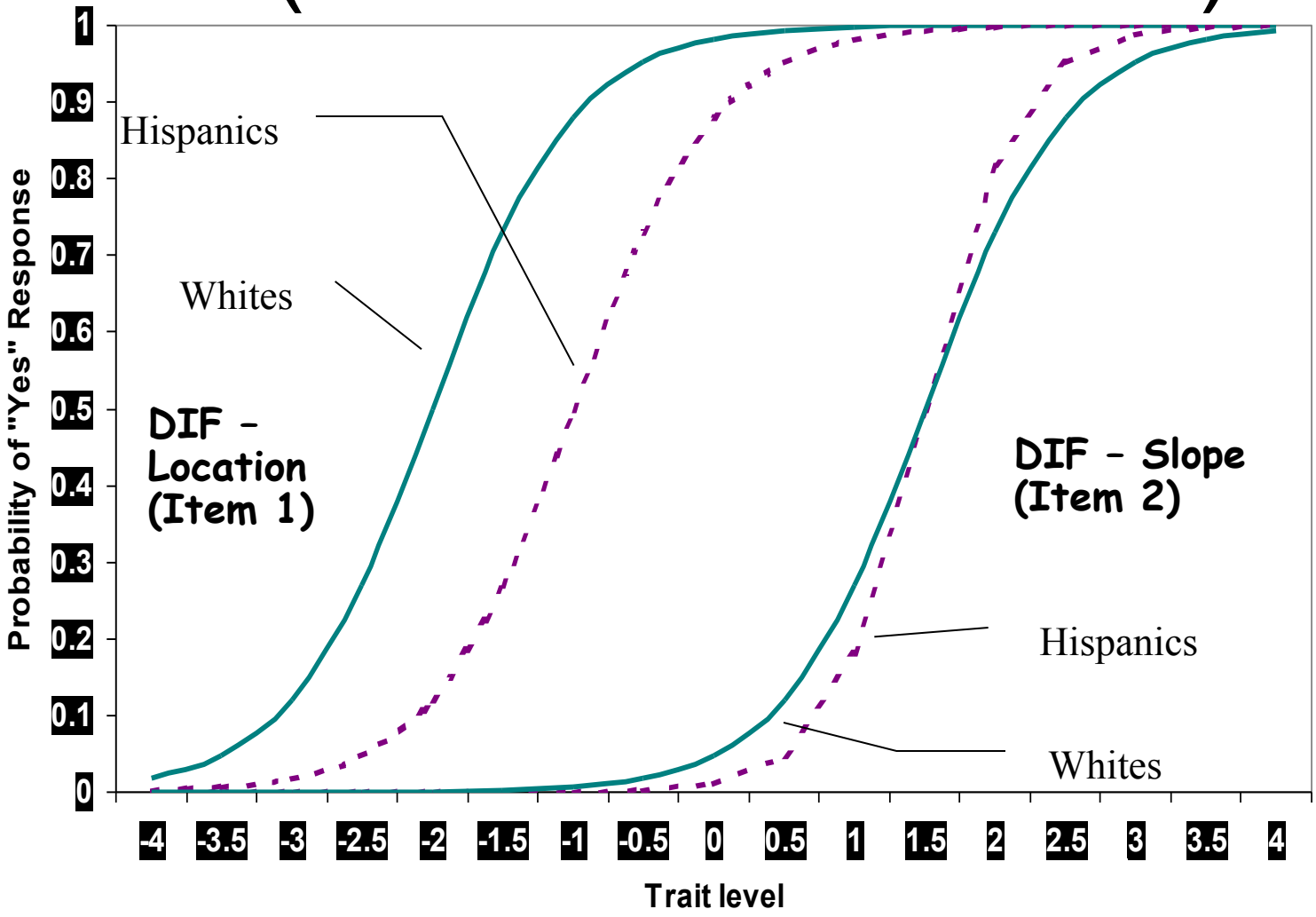
Item Characteristic Curves (1-Parameter Model)



Item Characteristic Curves (2-Parameter Model)



Dichotomous Items Showing DIF (2-Parameter Model)



Part IV: Steps to Develop a Measure and Residual Questions

Measurement Steps

- Review literature
- Expert input (patients and clinicians)
- Define constructs you are interested in
- Draft items (item generation)
- Pretest
  - Cognitive interviews
  - Field and pilot testing
- Revise and test again
- Translate/harmonize across languages

Residual Questions

drhays@ucla.edu
hays@rand.org