
Survey research: an overview of statistical and psychometric methods

Vanderbilt-Ingram Cancer Center Biostatistics Workshop
June 19, 2009

Irene D. Feuer, Ph.D.
Research Professor of Surgery and Biostatistics
Vanderbilt University Medical Center
irene.feurer@vanderbilt.edu

Workshop objectives

- **Discuss types of surveys and their purposes**
 - Surveys that are “scored”
 - Surveys requiring response summaries only
- **Discuss review criteria for survey instruments**
 - Scientific advisory committee of the Medical Outcomes Trust (Lohr et al. *Clinical Therapeutics* 1996;18:979-92.)
 - 8 criteria
- **Focus on classical test (true-score) theory (not IRT)**
- **Examples**

Examples

- **Summary response survey**
 - **Vanderbilt Transplant Center Employment survey (pre and post-transplant forms)**

- **Scored instruments**
 - **SF-36®**
 - **Vanderbilt Transplant Center Patient Satisfaction Inventory**
 - **EQ-5D**
 - **“Memphis” symptom survey**
 - **QDQ, Vanderbilt Bill Wilkerson Center for Otolaryngology Quantitative Dizziness Questionnaire (*GP Jacobson, et al., unpublished data*)**

Classical true-score theory: some assumptions

Error of measurement is unsystematic or random:

$$X = T + E$$

additivity of true score plus error of measurement

$$\xi(X) = T$$

expected value (population mean) of X is T

$$\rho_{ET} = 0$$

error and true scores from a population are uncorrelated

$$\rho_{E_1E_2} = 0$$

error scores of two different tests are uncorrelated

$$\rho_{E_1T_2} = 0$$

error scores of one test are uncorrelated with error scores of another test

MOT scientific review criteria and attributes

- **Conceptual and measurement model**
- **Reliability**
- **Validity**
- **Responsiveness**
- **Interpretability**
- **Respondent and administrative burden**
- **Alternative forms**
- **Cultural and language adaptations**

MOT scientific review criteria and attributes

Conceptual and measurement model

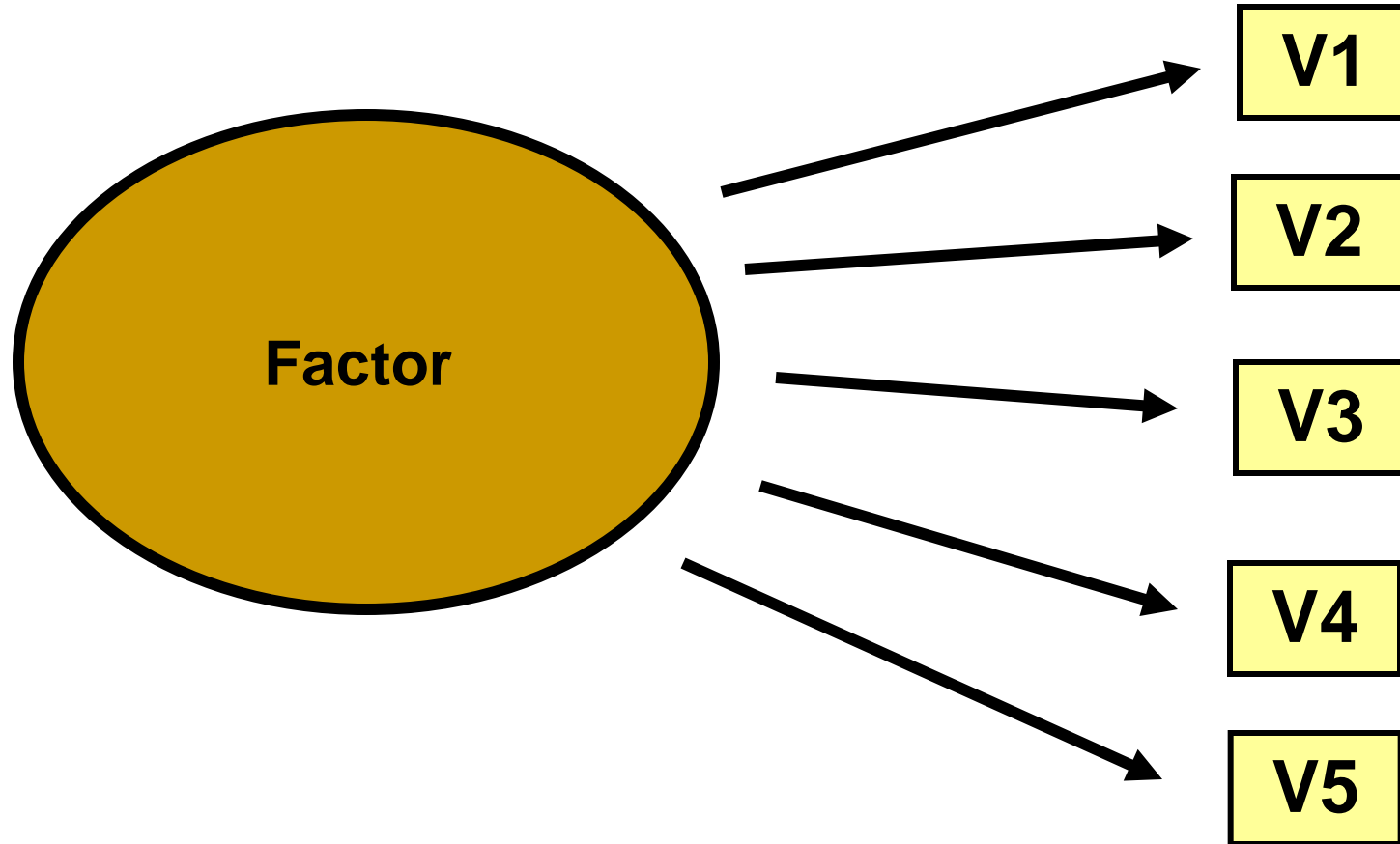
- Underlying rationale and description of concepts measured
 - Is reflected in survey's scale of measurement
 - Is reflected in subscale structure
1. What is the basis for combining items?
 2. What descriptive statistics are appropriate for scales?
 3. What evidence supports intended level of measurement (nominal, ordinal, interval, ratio scales)
 4. Are the procedures for deriving scales scores from raw scores justified?

MOT scientific review criteria and attributes

Reliability

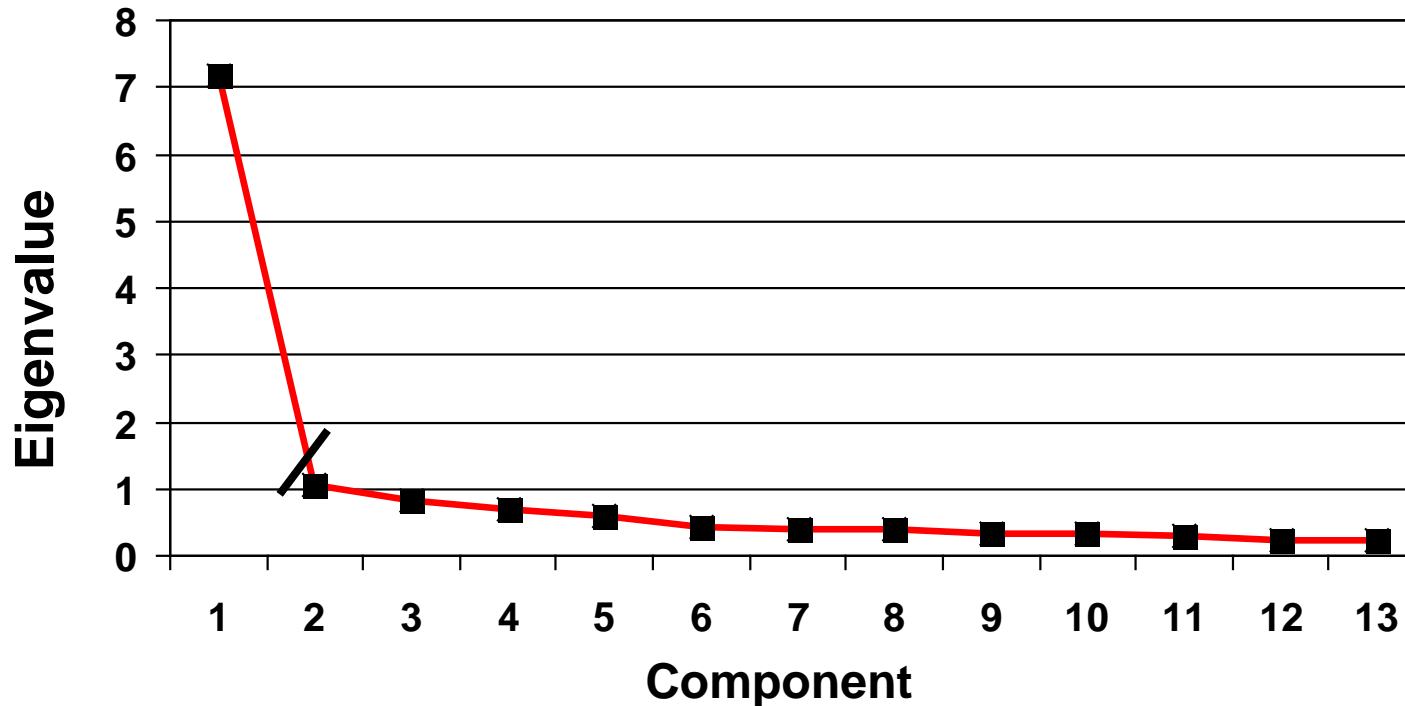
- Degree to which instrument is free of random (measurement) error
 - Reproducibility of scores under “equivalent” circumstances
1. Test-retest (paired test statistics)
 2. Inter-rater (kappa, ICC)
 3. Equivalent or alternate forms (r)
 4. Internal consistency (factor structure, coeff α)

Single latent variable (factor/component) with five manifest variables



Scree plot of eigenvalues

13 core Vanderbilt Transplant Center Patient Satisfaction Inventory (VTCPSI) items



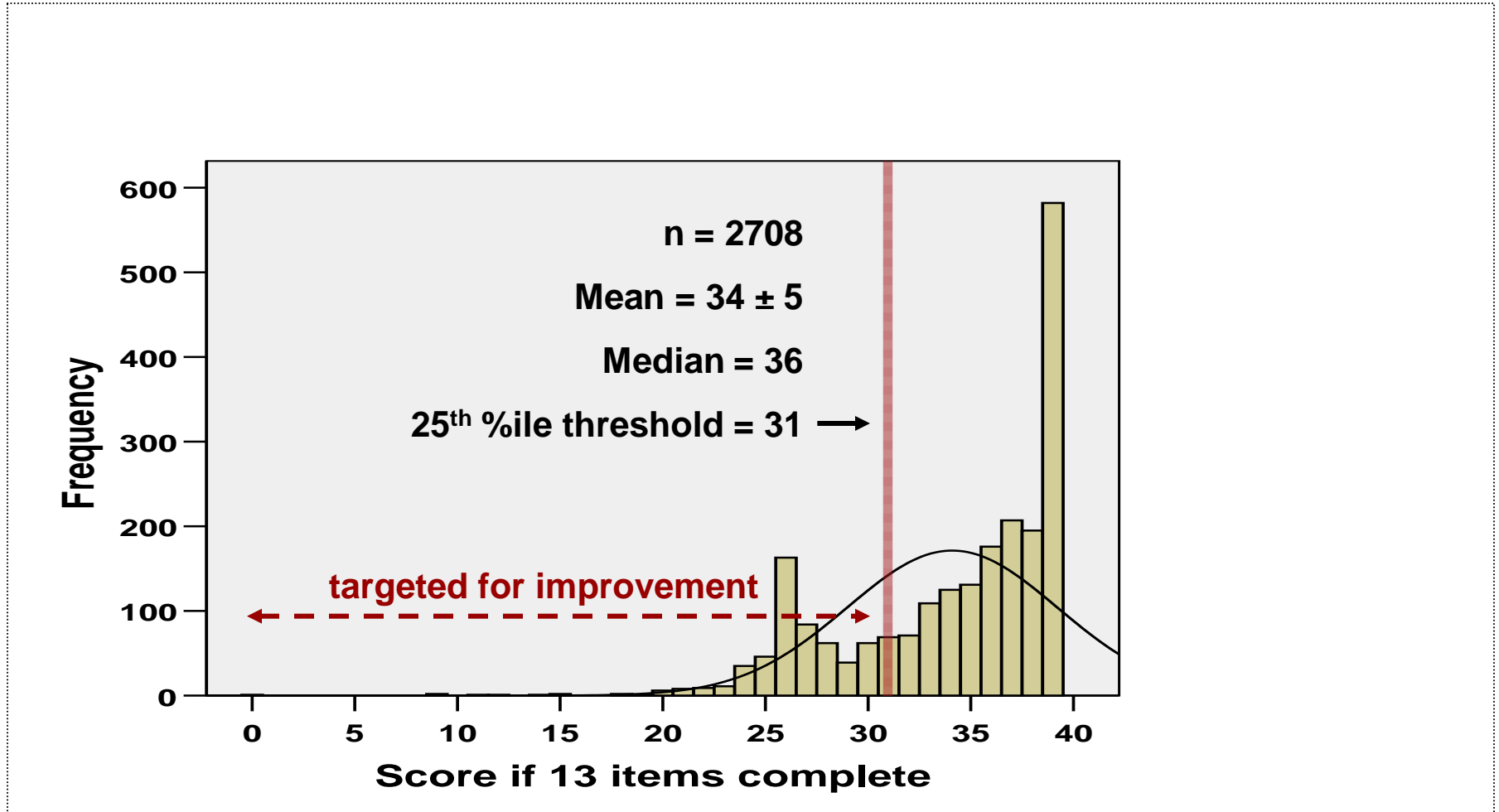
**The first component accounts for 56% of the covariance.
The second component accounts for only an additional 9%.**

Summary parameters - VTCPSI

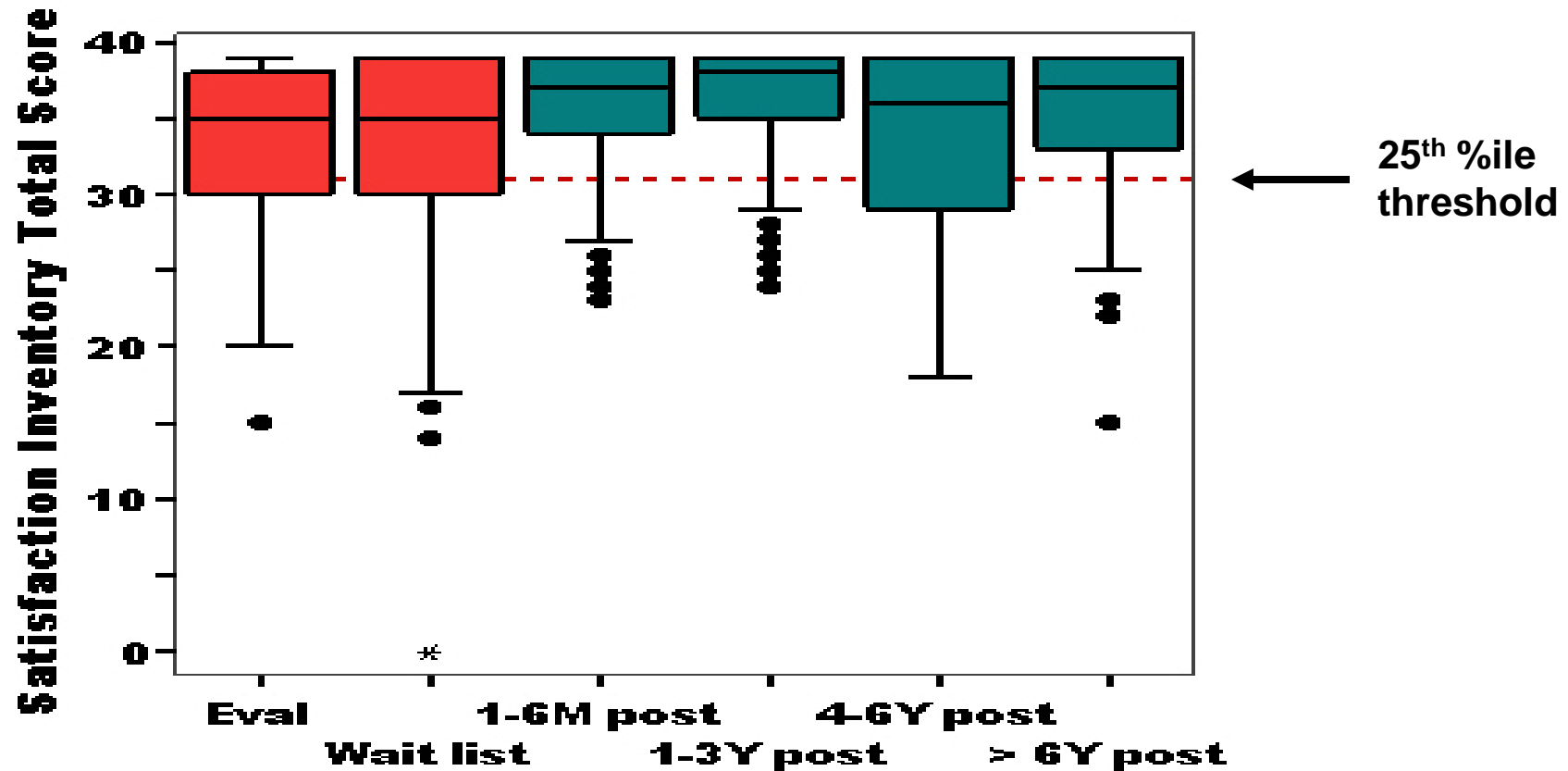
Time point	N	13 items complete	n:p	KMO	Eigen-val 1, 2	Comp 1 (% var)	Min load	Sat-uration	Coeff α
Eval	546	66%	28	.91	7.4 1.2	57%	.53	.75	.92
Wait	514	82%	32	.92	7.4 1.3	57%	.61	.75	.93
1 M	229	81%	14	.90	7.3 1.0	56%	.56	.74	.92
3 M	177	89%	12	.91	7.0 1.1	54%	.54	.73	.92
6 M	181	89%	12	.91	6.9 1.3	53%	.54	.72	.91
Ann	1061	86%	70	.95	7.2 1.2	56%	.51	.74	.92
Total	2708	81%	169	.94	7.2 1.2	56%	.55	.74	.92

Vanderbilt Transplant Center Patient Satisfaction Inventory[©]

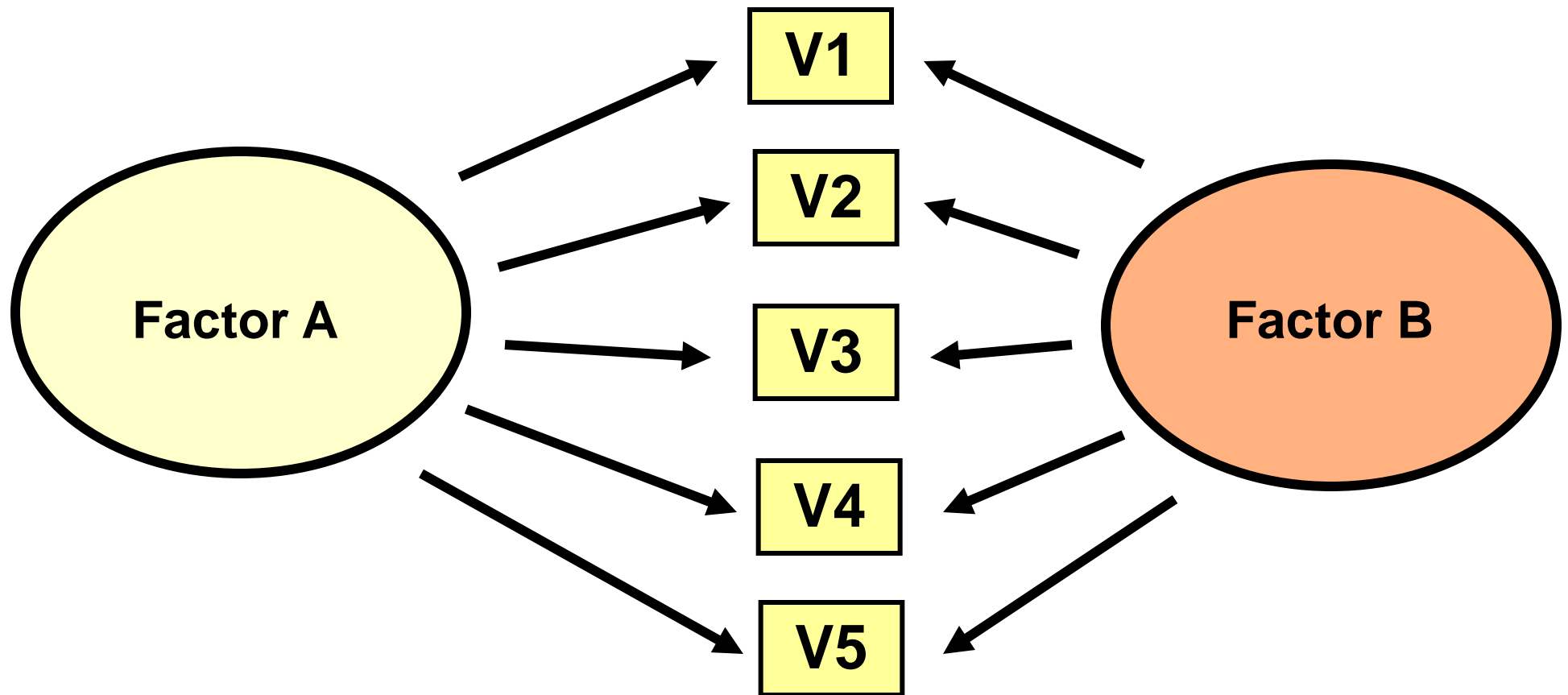
Score interpretation



Patient satisfaction with transplant healthcare delivery



Two latent variables and five manifest variables



SF-36® (v1) PCS and MCS

- **Eight subscales**

Physical function (PF)

Role physical (RP)

Bodily pain (BP)

General health (GH)

Vitality (VT)

Social functioning (SF)

Role emotional (RE)

Mental Health (MH)

- **Two summary components**

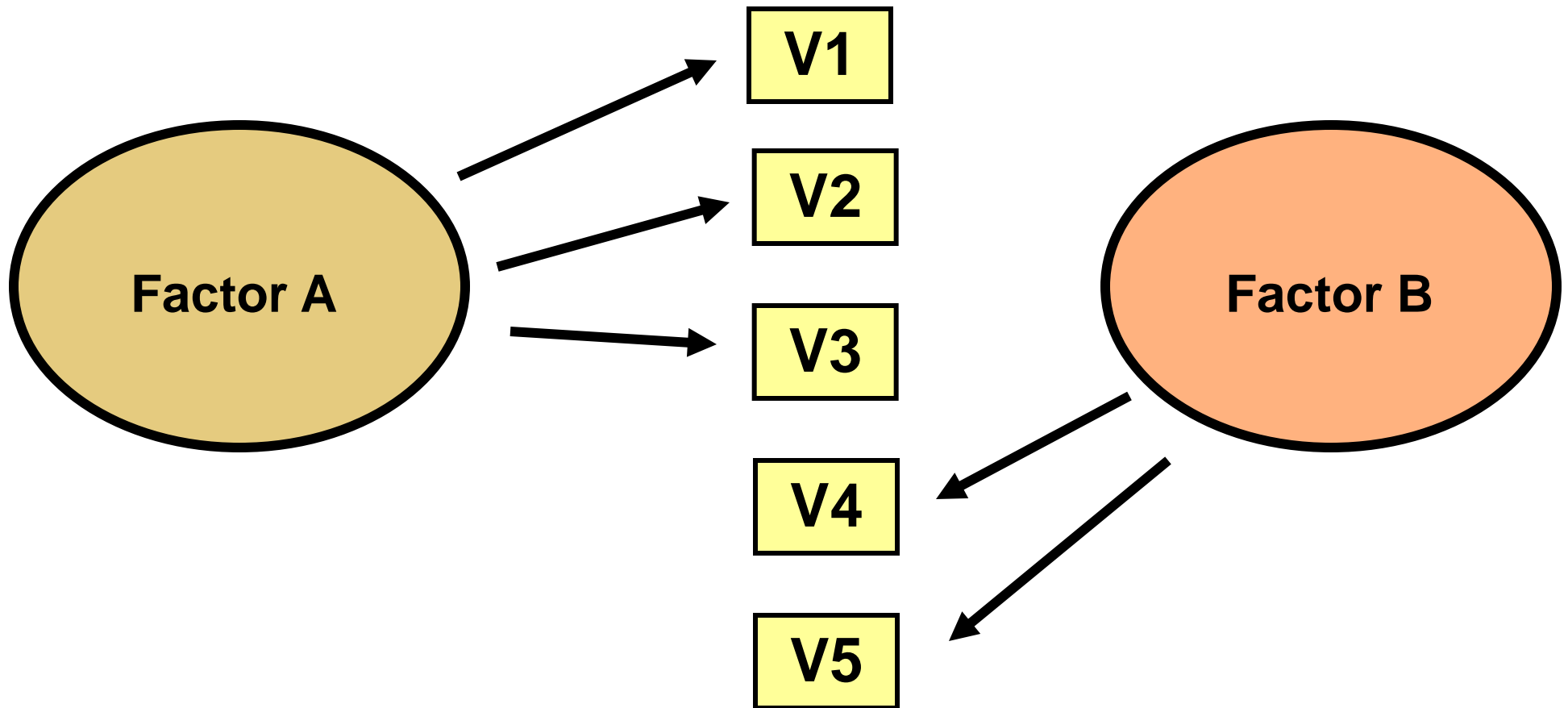
Physical (PCS)

Mental (MCS)

Scale (Z scores)	PCS	MCS
PF	.42402	-.22999
RP	.35119	-.12329
BP	.31754	-.09731
GH	.24954	-.01571
VT	.02887	.23534
SF	-.00753	.26876
RE	-.19206	.43407
MH	-.22069	.48581

Simple structure

two latent variables and five manifest variables



Quantitative Dizziness Questionnaire

Item analysis (unpublished data)

- Starting p = 43
- Starting n:p = 4.5 (n ~ 200)
- Final p = 33
- Content expertise – drop 2 items
- Component ambiguity – drop 4 items
- Internal consistency reliability (coefficient α) – drop 4 items
- 6 sales
 - Headache (5) $\alpha = .649$
 - Positional vertigo (6) $\alpha = .801$
 - Hydrops (5) $\alpha = .711$
 - Sup Canal Dehis (6) $\alpha = .863$
 - Multisensory (6) $\alpha = .791$
 - Psychophysiological (5) $\alpha = .836$

Unpublished data: Quantitative Dizziness Questionnaire (preliminary) headache scale

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.618	.615	6

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
G3	8.17	19.747	.124	.099	.649
G5	9.11	17.361	.292	.088	.597
H1	9.21	15.460	.449	.405	.532
H2	9.60	16.221	.454	.439	.535
H6	9.71	15.918	.328	.151	.587
H7	9.47	15.798	.469	.240	.527

MOT scientific review criteria and attributes

Validity

- Degree to which a survey is actually measuring the intended concept
 - A survey must be shown to be reliable *before* it can be considered valid
-
1. Content validity (related to “face validity”)
 2. Criterion-related validity (reference to gold standard)
 3. Construct validity (scores “behave as expected”)

HRQOL survey battery and assessment schedule

(Vanderbilt Transplant Center, effective January, 2002)

Survey Instrument	(# items)	Eval	Listed q 6 M	6 Hr Pre	1 M Post	3 M Post	6 M Post	Ann Post
SF-36®	(36)	X	X		X	X	X	X
Beck Anxiety Inventory	(21)	X	X		X	X	X	X
CES-D Depression Scale	(20)	X	X		X	X	X	X
Satisfaction Inventory©	(16)	X	X		X	X	X	X
Overall Health (VAS)	(1)	X	X		X	X	X	X
Employment	(20 pre, 17 post)	X					X	X
EQ-5D	(5)	X	X		X	X	X	X
Symptom survey	(pending)	X	X		X	X	X	X
Functional Performance	(1)	X	X	X	X	X	X	X

PAIS discontinued January, 2002

EQ-5D added July, 2006

↑
Listed

↑
Transplant

Criterion-related validity of EQ-5D preference weights in 285 liver transplant candidates and recipients theoretically-targeted associations

	PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH	BAI	CES-D
EQ-5D HUI	X	X										
Mobility	X		X	X		X	X					
Self-Care			X	X				X				
Usual Activities				X				X				
Pain	X				X							
Anxiety/Dep		X							X	X	X	X
VAS	X	X										

Dimensions and HUI of EQ-5D should have a strong correlation with scales/subscales from the “gold-standard” that measures similar constructs/domains (Spearman $r \geq 0.5$)

Criterion-Related Validity (with correction for “attenuation”)

- Classic test theory model
- Establishes the criterion-related validity of a target test (T_x) after adjusting for the known reliability of the target test and/or the gold standard (T_y)

$$r_{T_{EQ} T_{SF-36}} = r_{T_x T_y} = \frac{r_{xy}}{(r_{xx'} * r_{yy'})^{1/2}}$$

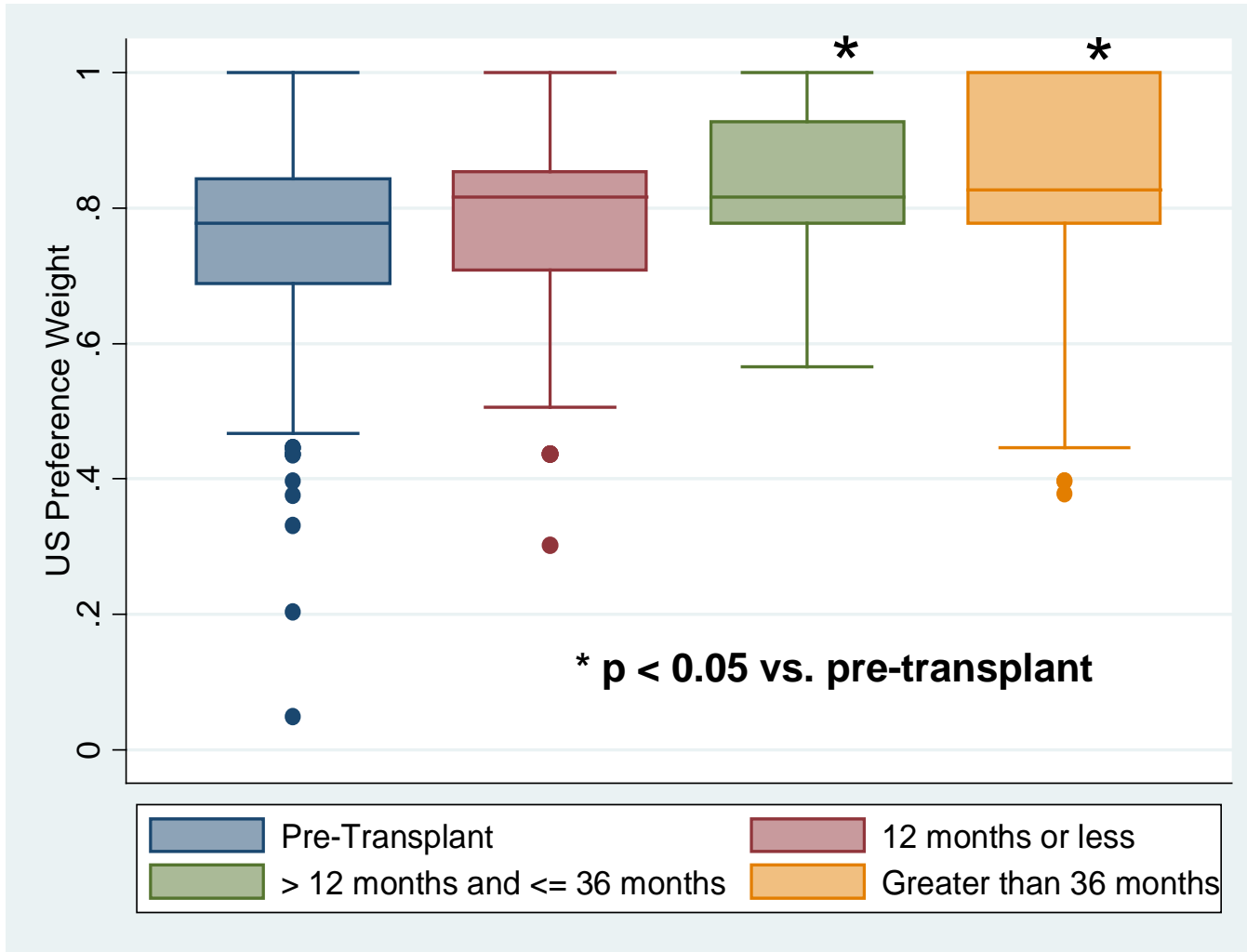
$r_{T_x T_y}$ = adjusted validity coefficient

r_{xy} = unadjusted correlation between T_x and T_y

$r_{xx'}$ and $r_{yy'}$ = reliability coefficients

(e.g. test-retest reliability from literature)

EQ-5D US population-based preference weights by period in liver transplant patients



EQ-5D US preference weights provide a valid metric for cost-utility analysis in the liver transplant setting.

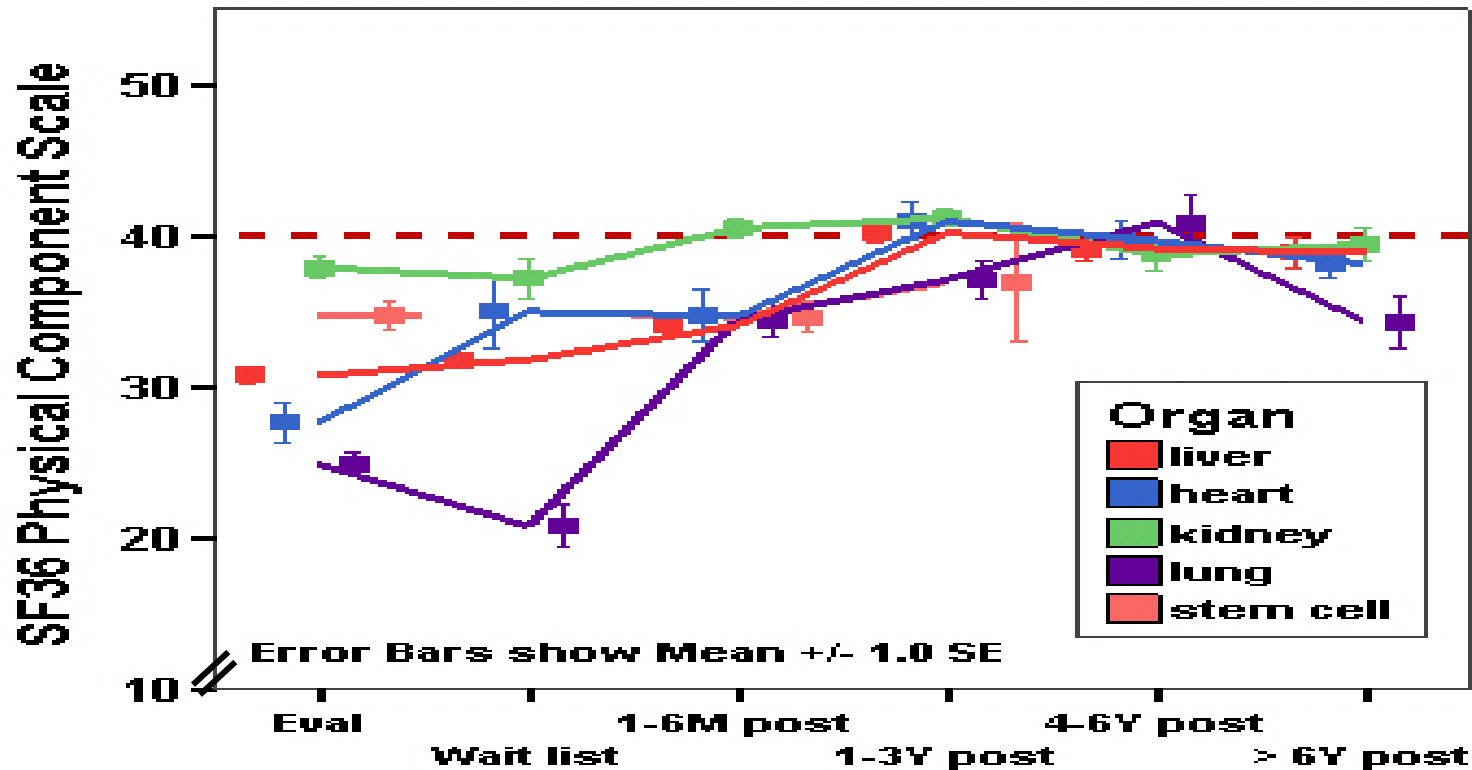
MOT scientific review criteria and attributes

Responsiveness

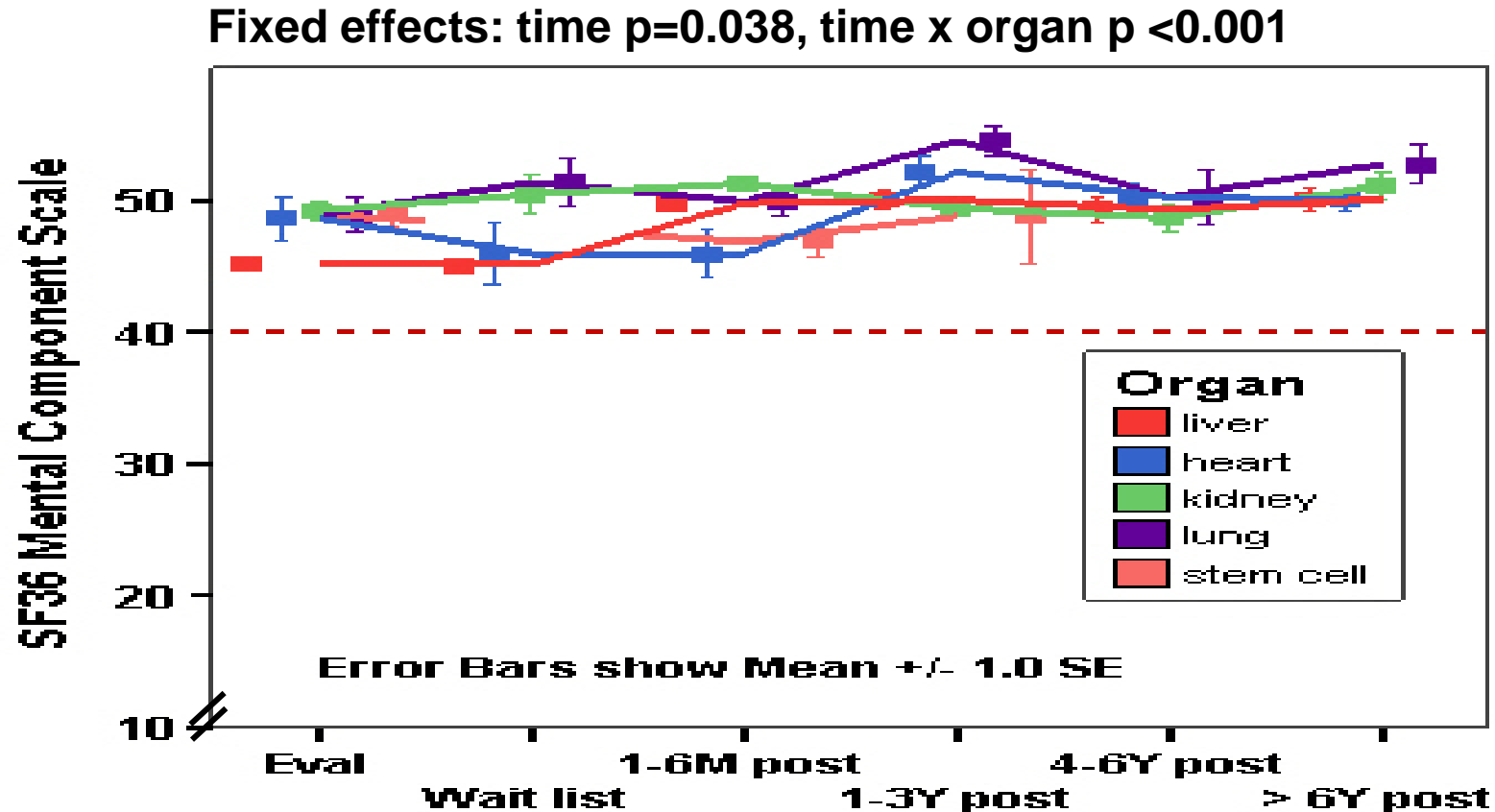
- **Instrument's ability to detect change over time**
 1. **longitudinal data**
 2. **effect sizes**
 3. **clinically relevant change**
 4. **“signal to noise ratio”**

Effect of time on physical HRQOL by transplant organ group

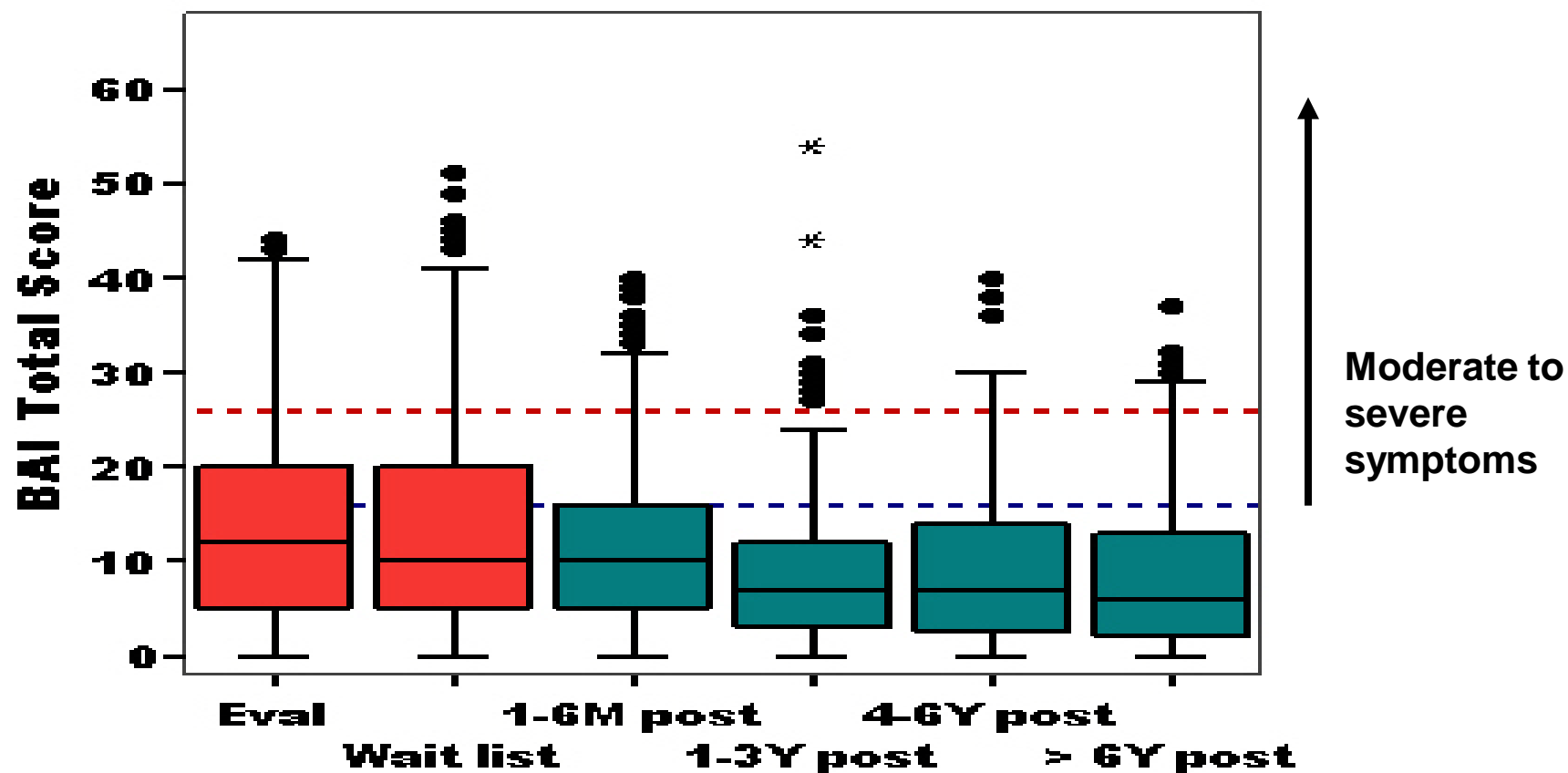
Fixed effects: time $p=0.001$, time x organ $p < 0.001$



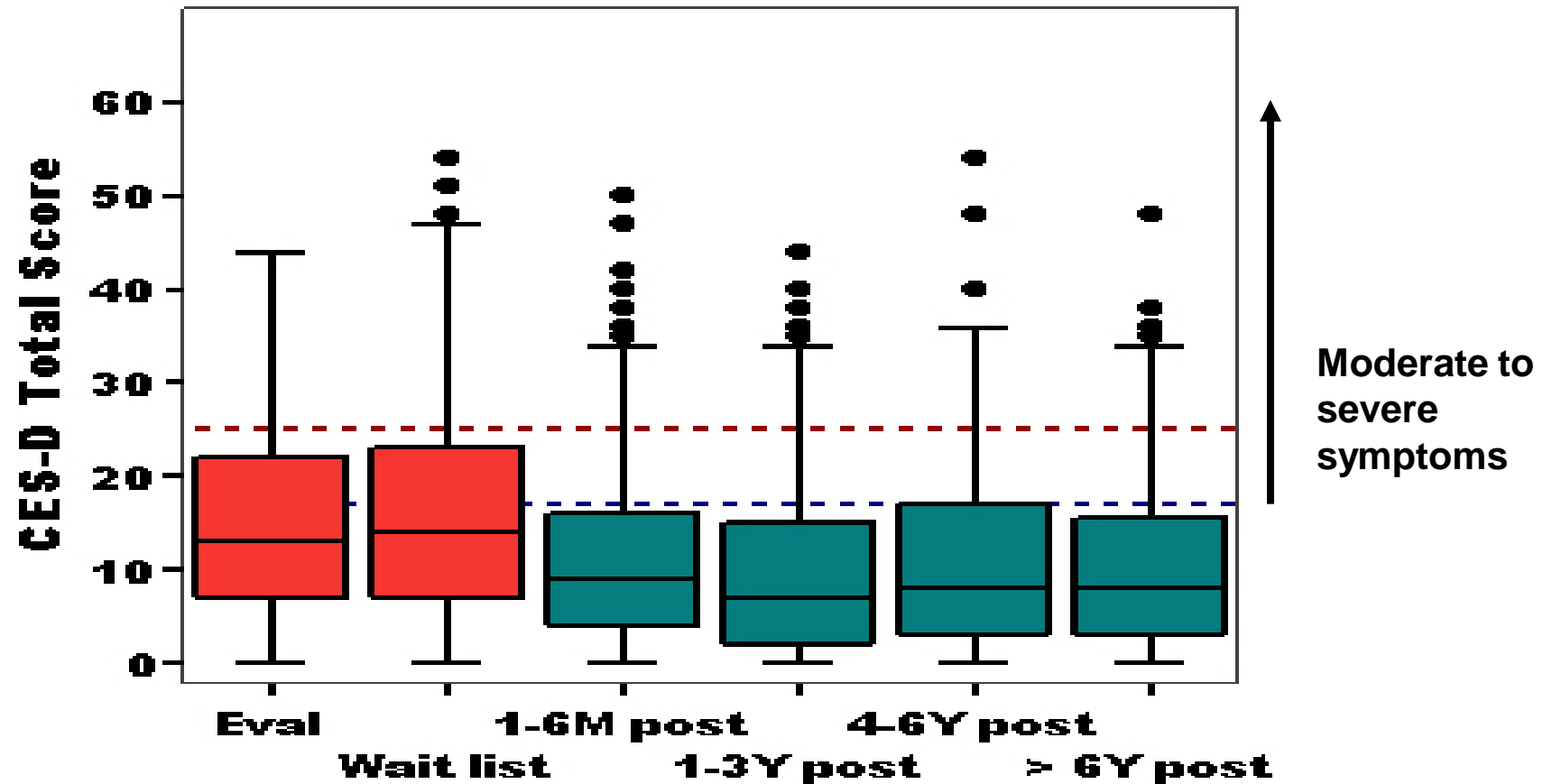
Effect of time on mental HRQOL by transplant organ group



Anxiety symptoms by measurement period



Depression symptoms by measurement period



MOT scientific review criteria and attributes

Respondent and administrative burden

- **Strike a balance between construct coverage and cost in terms of time**
- **Related to survey completion rates**

Original patient report survey battery redundancy of PAIS and SF-36 scales

PAIS → ↙ SF-36	Voc	Dom	Sex	Fam	Soc	Psy	HC	R
PCS	0.20	0.24	0.08		0.27			0.65 *
MCS				0.12	0.18	0.54		0.72 *

Entries are standardized regression coefficients (β)

* P < 0.001

HRQOL survey battery and assessment schedule (Vanderbilt Transplant Center, effective January, 2002)

Survey Instrument	(# items)	Eval	Listed q 6 M	6 Hr Pre	1 M Post	3 M Post	6 M Post	Ann Post
SF-36®	(36)	X	X		X	X	X	X
Beck Anxiety Inventory	(21)	X	X		X	X	X	X
CES-D Depression Scale	(20)	X	X		X	X	X	X
Satisfaction Inventory©	(16)	X	X		X	X	X	X
Overall Health (VAS)	(1)	X	X		X	X	X	X
Employment	(20 pre, 17 post)	X					X	X
EQ-5D	(5)	X	X		X	X	X	X
Symptom survey	(pending)	X	X		X	X	X	X
Functional Performance	(1)	X	X	X	X	X	X	X

PAIS discontinued January, 2002

EQ-5D added July, 2006

↑
Listed

↑
Transplant

MOT scientific review criteria and attributes

Alternate forms

Cultural and language adaptations

Interpretability

- Ability to administer instrument in another format
- Validated alternative language forms
- Can qualitative meaning be assigned to quantitative scores?

Additional considerations

- **Availability of suitable norms**
- **“Rules” for handling missing item data**

Developing and refining a survey battery striking a balance

- Validated set of surveys (literature, context, etc.)
- Should you develop a survey “from scratch”
- Response burden
- Breadth of constructs
- Relevance to the clinical setting or research questions
- Redundancy
- Ongoing process