DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

Responsiveness and interpretation – concepts and methods

Henrik Hein Lauridsen

Overview

What is responsiveness and interpretation?

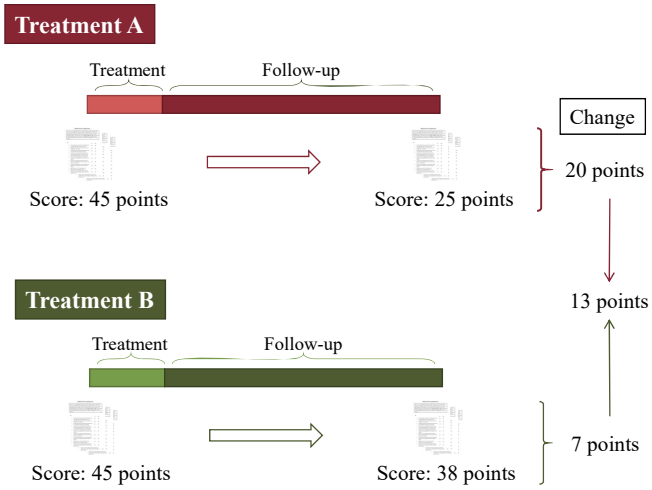
Responsiveness

- Construct responsiveness
- Criterion responsiveness

Interpretation

- Minimal Important Change (MIC)
- Anchor-based MIC distribution method

Measuring change



Is the measured change reproducible and valid?

How do we interpret the scores?

RESPONSIVENESS

Definition

Responsiveness

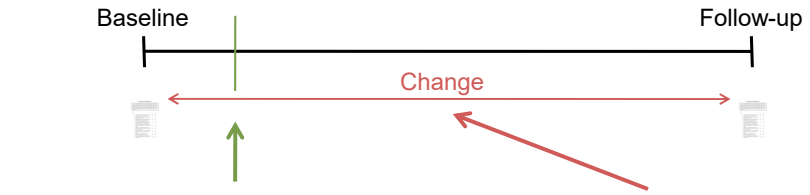
"The ability of an HR-PRO instrument to detect change over time in the construct to be measured"

COSMIN, 2011

"Målbarhed af en reel ændring"

Bendix T., 2008

Responsivness versus validity



Validity
"Cross-sectional" validity
Refers to validity of a single score

Responsivness
"Longitudinal" validity
Refers to validity of a score change

Consequently, validity and responsiveness have similar standards

Hays & Hardon, Qual Life Res 1992

How do we measure responsiveness?

31 different measures of responsiveness
6 different ways of grouping the measures

Typically reported as

- P-values
- Effect sizes (ES, SRM, Guyatt's Responsiveness statistics etc.)
- Correlations

Problems

- Many metodological problems → use with great care

Distribution- versus anchor-based approaches

Distribution-based methods

- Relates differences between baseline and follow-up scores to some measure of variability

Anchor-based methods

- Relationship between a questionnaire and an independent measure (or anchor)
- Elucidate the meaning of a particular degree of change

Distribution-based approaches

Effect size (ES)
- Individual change in SD of pretest

$$ES = \frac{Score_{pretest} - Score_{posttest}}{SD_{pretest}}$$

Standardized response mean (SRM)
- Individual change in SD of change

$$SRM = \frac{Score_{pretest} - Score_{posttest}}{SD_{change}}$$

Guyatt's responsiveness statistics (GRS)
- Accounts for spurious changes arising from measurement error

$$GRS = \frac{Minimal\ important\ change}{SD_{change,stable}}$$

$$GRS = \frac{(Score_{pretest} - Score_{posttest})_{Improved\ pts.}}{SD_{change,stable}}$$

Crosby et al. Defining clinically meaningful change in health-related quality of life. 2002

Problems with ES, SRM & GRS

Is not a reflection of the 'true change'
- Possible reasons:
  - Instrument could have a ceiling effect
  - Instrument lack relevant items

Highly dependent on variability (SD)
- High coefficients with smal SD

Are measures of the *magnitude* and not the *validity* of change scores

Two types of responsiveness

Construct responsiveness

The degree to which the change scores of a HR-PRO instrument are consistent with hypotheses

- Tests hypotheses about expected change

Criterion responsiveness

The degree to which the change scores of a HR-PRO instrument are an adequate reflection of a 'gold standard'

- Correlations
- Sensitivity/specificity
- Receiver Operating Characteristic (ROC) curves

Mokkink et al. (2010); Terwee et al. (2009)

Construct responsiveness

Specify the hypotheses before collecting data

Eg.:

- Expected correlations between score changes and changes in other variables
  - Example: function and satisfaction with treatment
- Expected differences in score changes between groups
  - Example: severe depression versus mild depression
- Expected score changes after a treatment with known effect
  - Example: treatment of tension headache with NSAID

Terwee et al. (2009)

Construct responsiveness

Hypothesis

- Should be specific
  - Direction of the expected correlation/difference
  - Size of the expected correlation/difference
- Should be challenging
- The more hypotheses the better

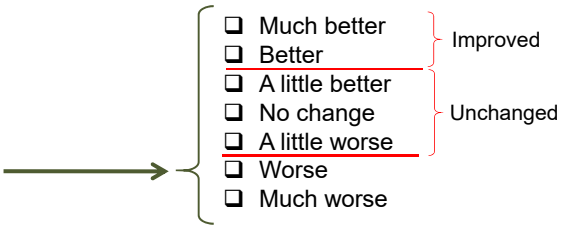
Terwee et al. (2009)

Construct responsiveness

”Low Vision Quality of Life” questionnaire (LVQoL)

4 subscales
- Basic aspects
- Mobility
- Adjustment
- Reading

Compared to
- VF-14
- Acuity
- Global change



De Boer et al. Qual Life Res 2006

Construct responsiveness

Hypotheses example

Correlation between score changes on LVQoL and global change is least 0.10 higher compared to the correlation between the score changes on LVQoL and acuity

Results

LVQoL subscale	Global change		Change in acuity
Δ basic aspects	0.26	>	0.13
Δ mobility	0.26	>	0.08
Δ reading	0.18	>	0.07
Δ adjustment	0.20	>	0.04

4 out of 4 (100%) hypotheses confirmed

De Boer et al. Qual Life Res 2006

Criterion responsiveness

Requires a 'gold standard' measuring change in the domain

Eg.

- Global change
- Change in a clinical variable

Methods

- Correlation between a 'gold standard' and score chagne
- Sensitivity/specificity – dichotomous variable
- Receiver Operating Characteristic (ROC) curves

ROC analysis

Treatment

Score: 25

Score: 18

Change
7 points

External anchor
(Retrospective)
"How are you now compared to when you started?"

- ❑ Much better
- ❑ Better
- ❑ A little better
- ❑ No change
- ❑ A little worse
- ❑ Worse
- ❑ Much worse

The external anchor (transition question)

"How are you now compared to when you started?"

Recall bias	Present state bias	Motivational bias	Contami-nation bias	Pleasing bias
Problems remembering prior healths status	Relate to how things are now and *not* to the change	Cumbersome treatment → Overestimation of the treatment response	Comorbidity contaminates the answer Eg. LBP patient with headache	Patients wants to please the clinician

Norman et al., 1997; Hägg et al., 2002; Middle et al., 2006; Aseltine et al. 1995; Herrmann et al., 1995, Lauridsen et al. 2007

Bias depends on time frame

Follow-up < approx. 6 wks.

- Retrospective anchor ≈ OK

Follow-up > approx. 6 wks.

- Retrospective anchor biases

Solution: 'The Punum Ladder'

The Punum Ladder

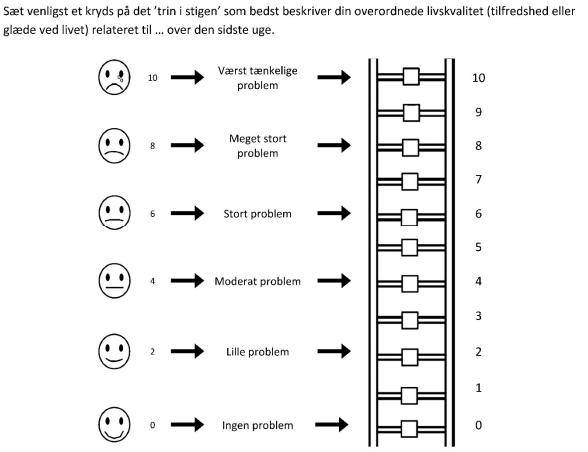
Prospective anchor which avoids biases

Score range: [0-10]

Applied to each dimension

Well validated:

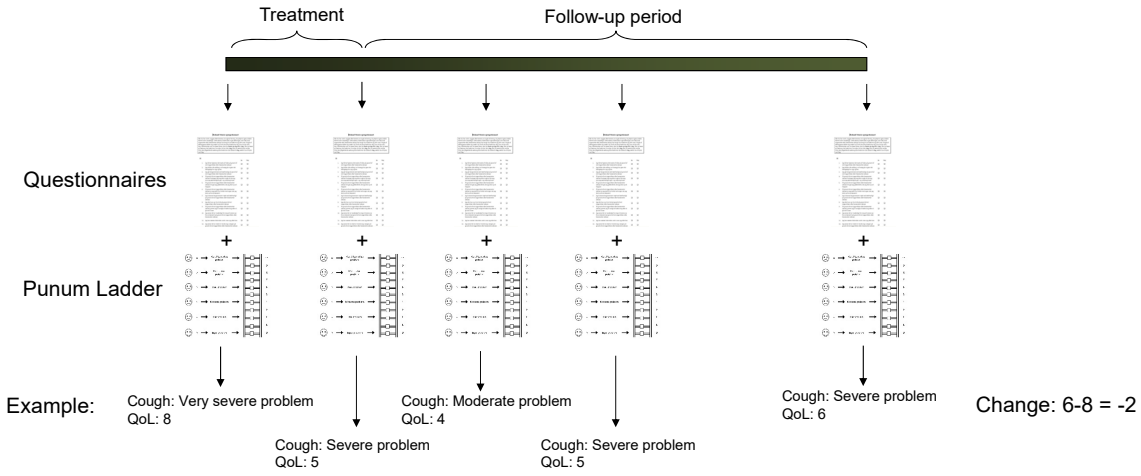
- Validity
- Reproducibility
  - Reliability
  - Internal consistency
- Responsiveness



NB: Danish version has not been tested for cross-cultural validity

Fletcher et al, 2010

Study design with Punum Ladder



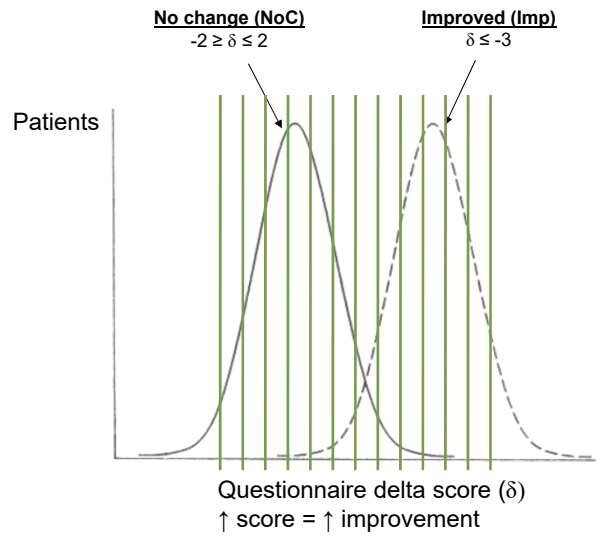
Fletcher et al, 2010

Dichotomisation

Change score ratings	Interpretation
-2, -1, 0, +1, +2	No change
≤ -3	Improvement
≥ +3	Deterioration

Fletcher et al, 2010

Sensitivity and specificity



		Anchor	
		Imp	NoC
Score	≥ δ	A	B
	< δ	C	D

Sensitivity: A/(A+C)

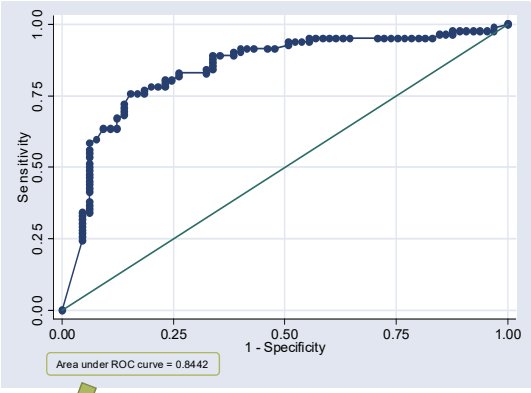
True positive rate

Specificity: D/(B+D)

True negative rate

The ROC curve

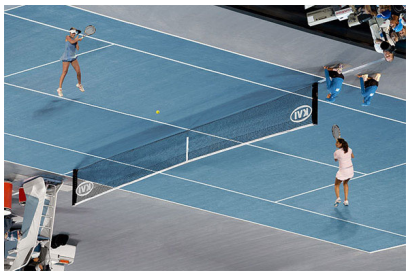
Plots the true positive rate (sensitivity) against the false positive rate (1-specificitet)



Area under the curve = proportion of correctly identified patients, who truly have improved = responsiveness

ROC_{auc} = 0.50 → unresponsive

Questions?



INTERPRETATION OF A SCORE CHANGE

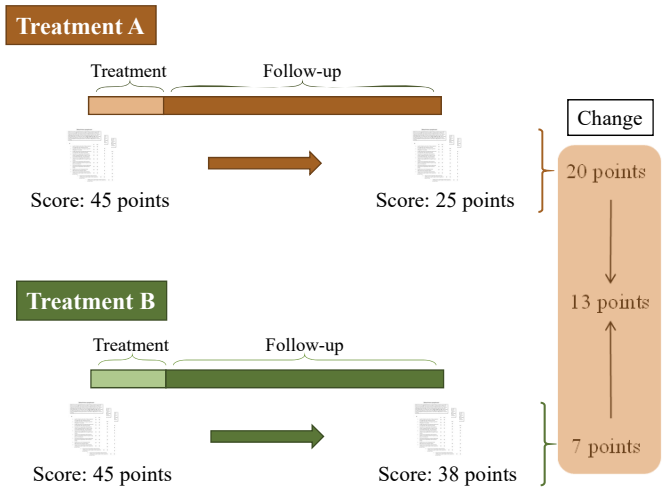
Interpretation

Definition

"The degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores"

Cosmin 2011

Interpretation



How do we interpret the scores?

Minimal Important Change (MIC)

Definition

"The smallest change in score in the construct to be measured which patients perceive as important"

Mokkink et al, 2010

Minimal Important Change

MID/MIC/MCID/MCIC

- MI = Minimal important
- MCI = Minimal clinically important
- D = difference
- C = change

Change: within persons/groups

Difference: between persons/groups

Two main approaches:

1. Anchor-based MIC distribution method
2. Predictive modelling method

van der Roer et al., Spine (2006); De Vet et al., Health Qual Life Out (2006); Terluin et al., J. Clin. Epid. (2015)

Anchor-based MIC distribution method

Graphical method to determine the best MIC interval

- Uses a globalt anchor:
  - Transition question method – retrospective
  - Punum Ladder method - prospective

Can be made for an 'important improvement' and 'important deterioration'

5-step process

de Vet et al. (2007)

SDU

Step 1-2

1. Choose a global anchor (transition question)
2. Define an 'important change'

Transition question method

"How are you now compared to when you started?"

- ❑ Much better
- ❑ Better

Important improvement

- ❑ A little better
- ❑ No change
- ❑ A little worse

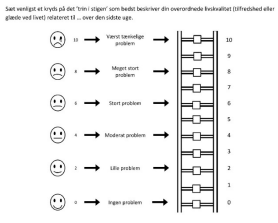
Unchanged

- ❑ Worse
- ❑ Much worse

Important deterioration

Punum Ladder method

"Change in QoL in the measured domain"



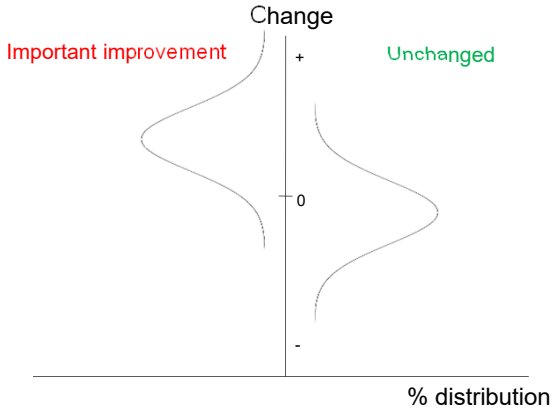
≤ -3 = Important improvement

-2, -1, 0, +1, +2 = Unchanged

SDU

Step 3

3. Draw the curves



SDU

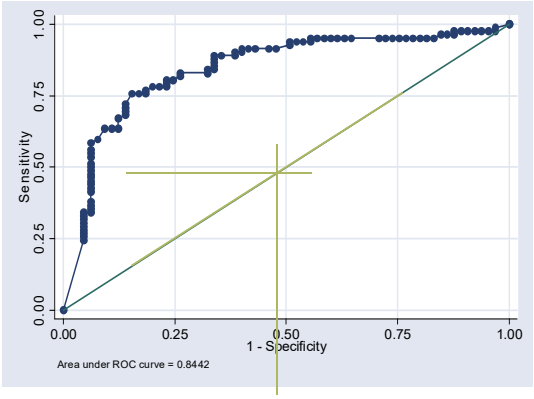
Step 4

4. Calc. MIC from ROC analysis

Point in the top left corner:

- Highest true positive rate
- Lowest false positive rate

Minimal Important Change



DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

Step 4

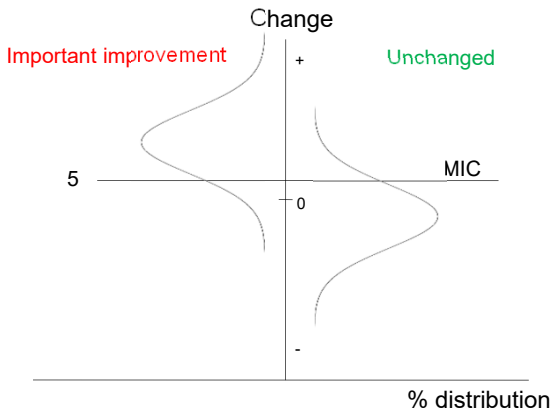
4. Calc. MIC – alternative method

- Calculate mean value of the domain in the Punum Ladder score ratings = -4, -3, +3, +4
- Minimal Important Change

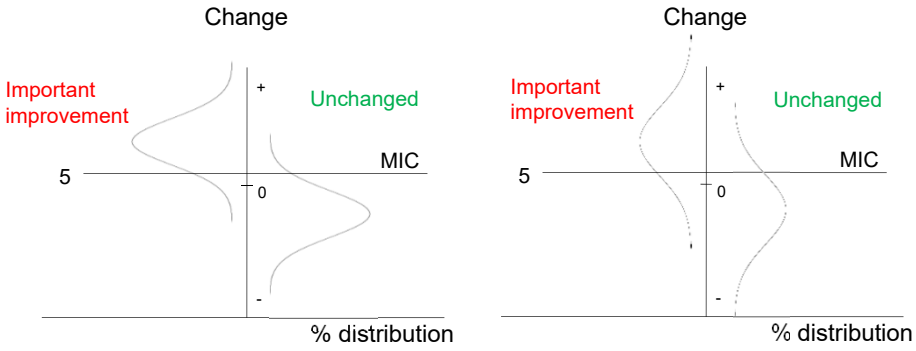
Fletcher et al, 2010

Step 5

5. Insert MIC in the graph

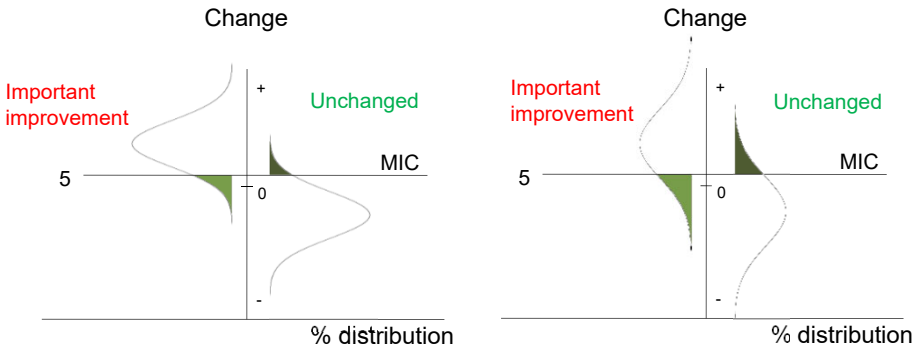


Interpretation of the graphs



Assess the hight of the curves:
Peaked curves → high correlation between anchor and score → High discrimination of the anchor

Interpretation of the graphs



- Assess misclassification of the curves
  - = false positive (1 – specificity) according to cut-point
  - = false negative (1 – sensitivity) according to cut-point

Questions?

SDU