HENRIK HEIN LAURIDSEN

Responsiveness and interpretation – concepts and methods
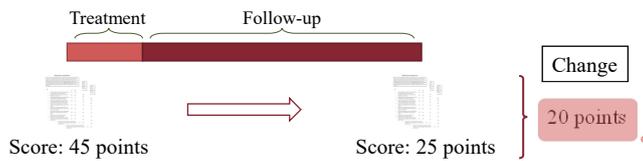
SDU

1

Measuring issues

Treatment A

Treatment   Follow-up

Score: 45 points   Score: 25 points
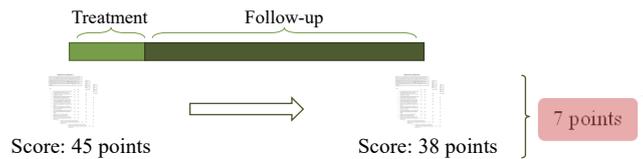
Change

20 points

Responsiveness
Can the questionnaire(s) measure the change accurately i.e. validly and reliably?

Treatment B

Treatment   Follow-up

Score: 45 points   Score: 38 points

7 points

SDU

2

1

Measuring issues

Treatment   Follow-up

Score: 45 points   Score: 25 points
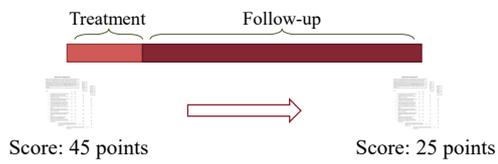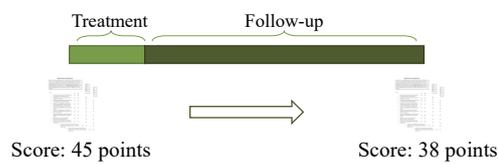
Change

20 points

13 points

7 points

Interpretation
How do we interpret the before/after and between group change scores?

Treatment B

Treatment   Follow-up

Score: 45 points   Score: 38 points

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

3

RESPONSIVENESS

SDU

4

Definition

Responsiveness

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

COSMIN, 2011

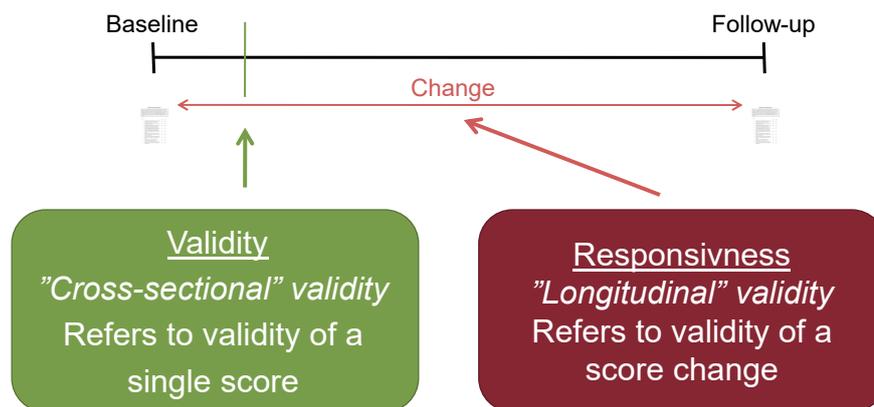*"Målbarhed af en reel ændring"*

Bendix T., 2008

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS SDU❧

5

Responsivness versus validity

Consequently, validity and responsiveness have similar standards

Hays & Hardon, Qual Life Res 1992

SDU❧

6

How do we measure responsiveness?

Numerous different measures of responsiveness
At least 6 different ways of grouping the methods

E.g.:
Internal & external responsiveness
Relative responsiveness
Sensitivity to change
Individual & group responsiveness
Distribution & Anchor-based responsiveness

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

Problems
- Many metodological problems → use with great care

SDU

7

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)
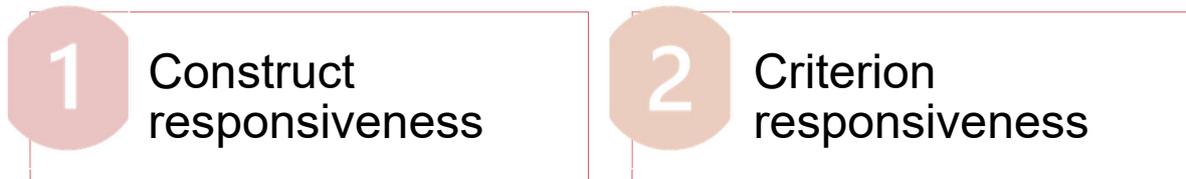- Elucidate the meaning of a particular degree of change

Advanced course

SDU

8

4

Two types of responsiveness

| 1 | Construct responsiveness |
| 2 | Criterion responsiveness |

SDU

9

Construct responsiveness

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

Tests hypotheses about expected change

E.g.
- Expected correlations between score changes and changes in other variables

- Example: function and satisfaction with treatment

Mokkink et al. (2018); Terwee et al. (2009)

SDU

10

Criterion responsiveness

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

Requires a 'gold standard' measuring change in the domain

E.g.
- Global change
- Change in a clinical variable

Mokkink et al. (2018); Terwee et al. (2009)

SDU

11

Criterion responsiveness

Level of measurement		Same units	Statistical parameter
Gold standard	Measurement instrument		
Dichotomous	Dichotomous	Yes	Sensitivity and specificity
	Ordinal	NA	ROC
	Continuous	NA	ROC
Ordinal	Ordinal	Yes	Weighted kappa
		No	Spearman's r^a or other measures of association
	Continuous	NA	ROCs[b]/Spearman's r
Continuous	Continuous	Yes	Bland and Altman limits of agreement or ICC[c]
		No	Spearman's r or Pearson's r

Advanced course

SDU

12

Criterion responsiveness with B & A LOA

> Both measure motor function of the affected arm in stroke patients

Purpose
- To compare the responsiveness of the ARA test and FMA scale

Responsiveness
- Compared responsiveness using B & A limits of agreement plots

J Rehab Med 2001; 33: 110–113

THE RESPONSIVENESS OF THE ACTION RESEARCH ARM TEST AND THE FUGL-MEYER ASSESSMENT SCALE IN CHRONIC STROKE PATIENTS

Johanna H. van der Lee,[1,2] Heleen Beckerman,[1,2] Gustaaf J. Lankhorst[1,2] and Lex M. Bouter[2]
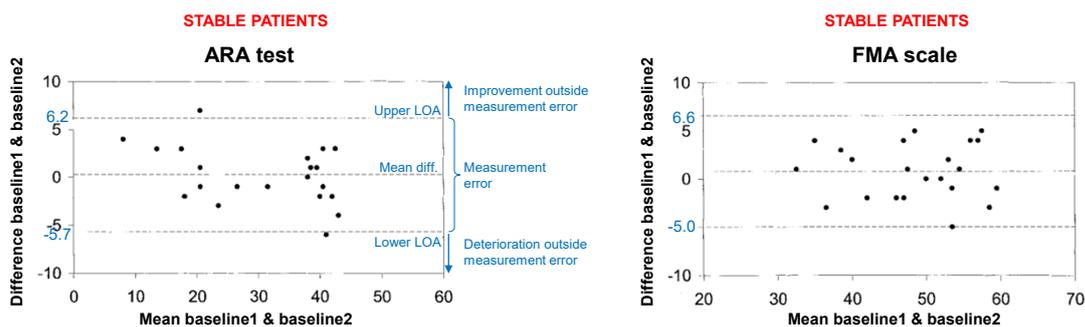
From the [1]Department of Rehabilitation Medicine, University Hospital Vrije Universiteit, [2]Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam, The Netherlands

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

13

Criterion responsiveness with B & A LOA

In an intervention group, how many improved outside measurement error?

Instrument	N	%
ARA test	12	54.5
FMA scale	2	9.1

Conclusion: The ARA test is more responsive

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

14

Criterion responsiveness with correlation

Both measure ADL
in stroke patients

Purpose

- To compare a new instrument (LIMOS) with an existing instrument (FIM), which has been shown to be responsive.

Responsiveness

- *Internal responsiveness:* The ability of a measure to change over a specific time frame

- *External responsiveness:* The extent to which changes in a measure relate to corresponding changes in a reference measure

This is criterion
responsiveness

frontiers
in Neurology

ORIGINAL RESEARCH
published: 26 September 2016
doi: 10.3389/fneur.2016.00152

The Responsiveness of the Lucerne ICF-Based Multidisciplinary Observation Scale: A Comparison with the Functional Independence Measure and the Barthel Index

Tim Vanbellingen[1,2]*, Beatrice Ottiger[1], Tobias Pflugshaupt[1], Jan Mehrholz[3], Stephan Bohlhalter[1], Tobias Nef[2,4] and Thomas Nyffeler[1,2]

15

Criterion responsiveness with correlation

Correlation with regression plot

Existing scale
with known
responsiveness

FIM motor subscale change score

LIMOS motor subscale change score

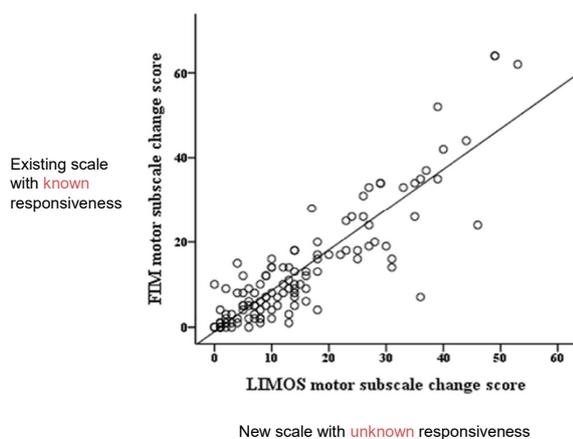New scale with unknown responsiveness

TABLE 3 | Measures of external responsiveness (correlation method and linear regression analysis) for the LIMOS motor and LIMOS cognition and communication with FIM motor and cognition.

Measures	r	p-value	β (SE)	p-value	R^2
LIMOS motor	0.88	<0.0001	0.73 (0.04)	<0.0001	0.84
LIMOS cog and commun	0.67	<0.0001	1.63 (0.18)	<0.0001	0.49

High correlation which is significant

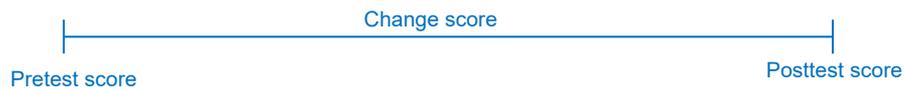For every 1 point change in LIMOS-motor we see a 0.73 point change in FIM.

The variation in the FIM-motor change explained 84% of the LIMOS-motor change.

16

Why are effect size not a good measure of responsiveness?

Groups of 2 x 2
Formulate an answer to the question

Change score

Pretest score   Posttest score

$$Effect\ size = \frac{Score_{pretest} - Score_{posttest}}{SD_{pretest}}$$

$$Standardiz\ Response\ Mean = \frac{Score_{pretest} - Score_{posttest}}{SD_{change}}$$

17

Questions?

18

Interpretation of change scores
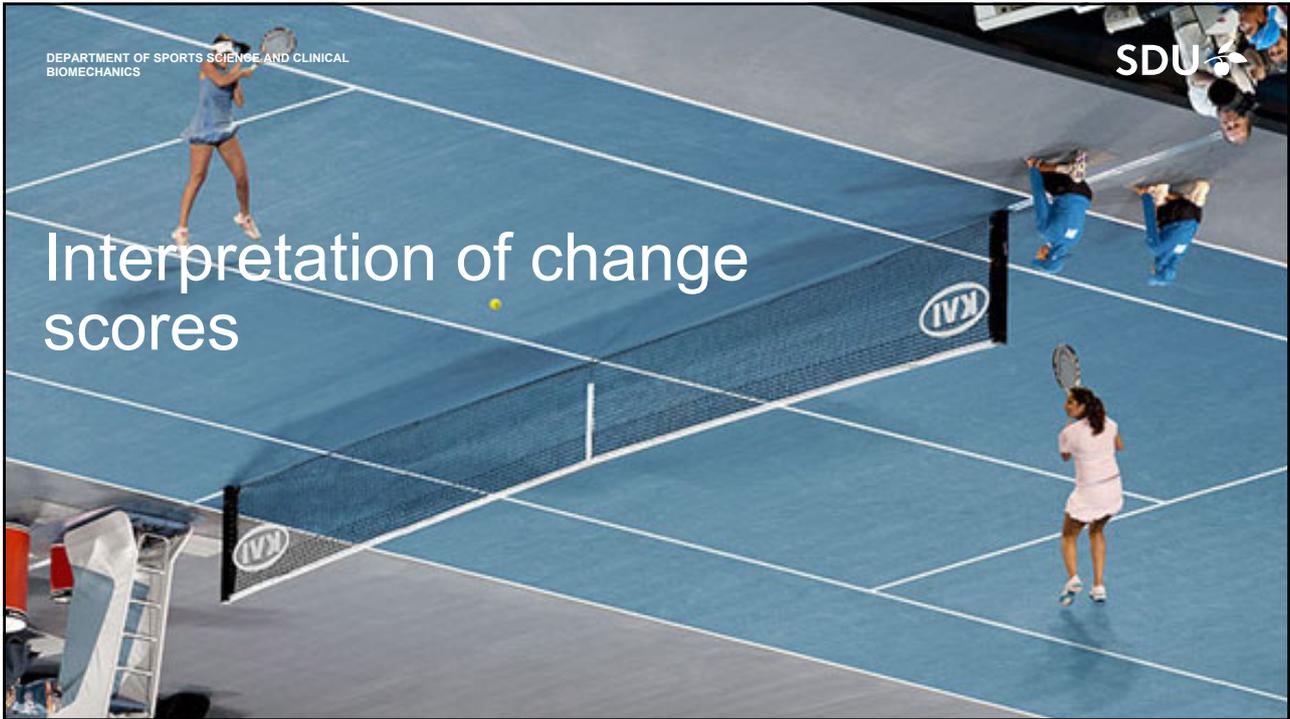
19

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

20

10

Minimal Important Change (MIC)

Mokkink et al. (2010)

Definition

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

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

21

Minimal Important Change

MID/MIC/MCID/MCIC

MI = Minimal important
MCI = Minimal clinically important
D = difference
C = change
… and many more

MID used by the Guyatt group (US)

MIC used by COSMIN

NO CONSENSUS

COSMIN

Change: within-patient or within-group change
Difference: between-patients or between-groups change
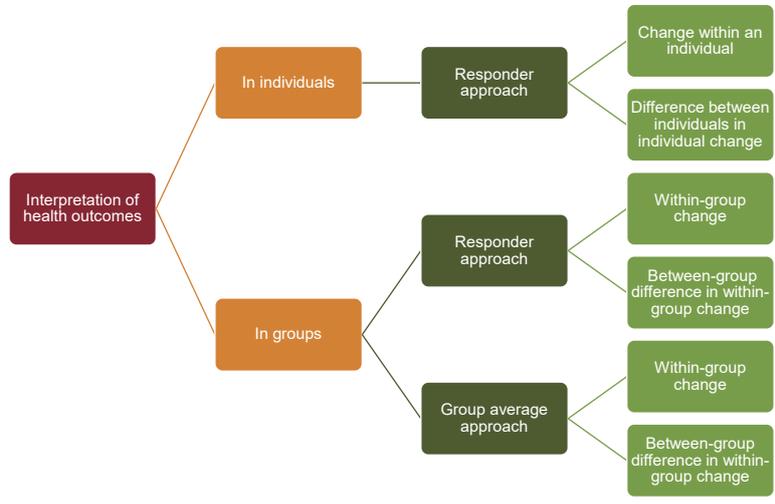
SDU

Guyatt et al., Mayo Clin Proc (2002); van der Roer et al., Spine (2006); De Vet et al., Health Qual Life Out (2006); Terluin et al., J. Clin. Epid. (2015); Dekker et al., Osteoarthritis and Cartilage (2023)

22

A conceptual framework

Dekker et al., Osteoarthritis and Cartilage (2023)

23

MIC and difference - methods

Change in…	Approach	Between who…	Concept	Methods	Criteria
Individuals	*Responder*	Change *within* individuals	Responder	MIC: • Distribution-based • Anchor-based • Consensus-based	$\Delta T_1 - T_0 > \text{MIC}$
				MIC$_{individual}$: • Conversation with individuals	$\Delta T_1 - T_0 > \text{individual MIC}$
				Responder criteria	Individuals meets responder criteria
				Goal attainment	Individual achieves goal
		Difference *between* changes within individuals	Tbd	Tbd	Tbd

Tbd = To be determined, NNT = numbers needed to treat, RR = relative risk, SWE = Smallest worthwhile effect, ΔA = within-group change

Dekker et al., Osteoarthritis and Cartilage (2023)

24

MIC and difference - methods

Change in…	Approach	Between who…	Concept	Methods	Criteria
Individuals	*Responder*	Change *within* individuals	Responder	MIC: • Distribution-based • Anchor-based • Consensus-based	$\Delta T_1 - T_0 > MIC$
				MIC$_{individual}$: • Conversation with individuals	$\Delta T_1 - T_0 >$ individual MIC
				Responder criteria	Individuals meets responder criteria
				Goal attainment	Individual achieves goal
		Difference *between* changes within individuals	Tbd	Tbd	Tbd
Groups	*Responder*	*Within-group* change	Proportion of responders	Consensus-based cut-off	Proportion of responders (%ΔA) > cut-off
		Between-group difference in within-group change		Consensus-based cut-off for maximal NNT or minimal RR	NNT < cut-off RR > cut-off
	Group average	*Within-group* change	Average change	MIC	Within-group change ΔA > MIC
		Between-group difference in within-group change		Anchor-based or consensus-based → MCID	Between-group diff. > MCID
				Benefit-harm trade-off or discrete choice experiment → SWE	Between-group diff. > SWE

Tbd = To be determined, NNT = numbers needed to treat, RR = relative risk, SWE = Smallest worthwhile effect, ΔA = within-group change

Dekker, Joost, Michiel de Boer, and Raymond Ostelo. 'Minimal Important Change and Difference in Health Outcome: An Overview of Approaches, Concepts, and Methods'. *Osteoarthritis and Cartilage*, 2023.

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

25

MIC calculation techniques

Distribution-based
- Based on statistical characteristics in the sample
- 3 major categories:
    - Statistical significance (e.g. paired t-statistics)
    - Sample variation (e.g. $\Delta T_1 - T_0$ = small effect size (0.2))
    - Measurement precision (standard error of the meaurement, SEM)

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

Anchor-based
- Based on another measure of change – external anchor
    - Mean group techniques
    - Receiver Operating Characteristics curves
    - Contemporary techniques
        - Predictive modelling (PM)
        - Adjusted Predictive Modelling (APM)
        - Improved Adjusted Predictive Modelling (IAPM)
        - Longitudinal Confirmatory Factor Analysis

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

26

The Global Rating of Change (GRC scale)

Also called external anchor or transition question/rating

Example:

Compared with before receiving treatment, how would you describe your physical abilities now?

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

Kamper SJ et al. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J.Man.Manip.Ther. 2009

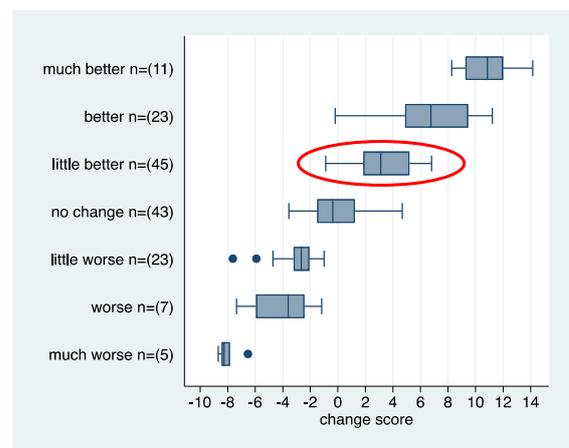DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

27

Most common techniques (anchor)

Mean value of the change score in patients reporting a minimal improvement

Most popular method

Carrasco-Labra A et al. Minimal important difference estimates for patient-reported outcomes: A systematic survey. J Clin Epidemiol. 2021

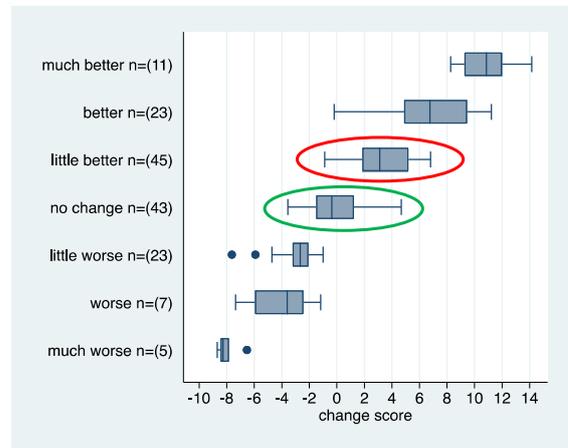DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

28

Most common techniques (anchor)

The difference in means value of the change score between patients reporting a minimal improvement and patients reporting no change

Third most popular method

much better n=(11)
better n=(23)
little better n=(45)
no change n=(43)
little worse n=(23)
worse n=(7)
much worse n=(5)

-10 -8 -6 -4 -2 0 2 4 6 8 10 12 14
change score

Carrasco-Labra A et al. Minimal important difference estimates for patient-reported outcomes: A systematic survey. J Clin Epidemiol. 2021

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS
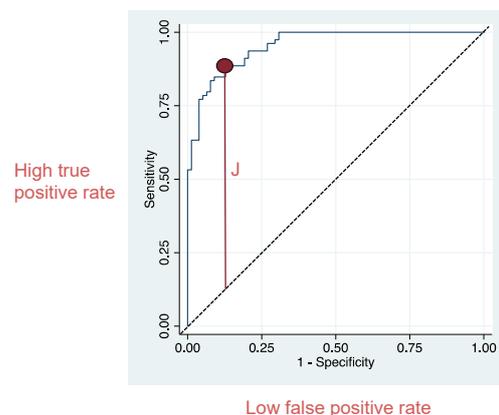
SDU

29

Most common techniques (anchor)

Receiver Operating Characteristic curve

Defines the MIC as the optimal cut-point in a diagnostic test

Different ways to define optimal cut-point e.g. Youden index (J)
- Equal weight to sensitivity and specificity
- Number between 0-1
- Highest value = MIC

Second most popular method

High true positive rate

J

Sensitivity

1 - Specificity

Low false positive rate

Carrasco-Labra A et al. Minimal important difference estimates for patient-reported outcomes: A systematic survey. J Clin Epidemiol. 2021

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

30

What can the MIC be used for?

- Responder analysis in clinical trials? **YES**

- Power analysis? **YES**

- Interpretation of longitudinal changes scores in groups? **YES**

- Interpretation of longitudinal changes scores in individuals? **Maybe if reliability is high enough**

- Interpretation of differences between groups in clinical trials? **NO (but debatable…)**

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

31

MIC details, calculation methods and howto!

Advanced course in clinimetrics and questionnaire technique

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

32

Main points – responsiveness & interpretation

Responsiveness
- Is the same as longitudinal validity
- Distribution-based coefficients are still used (e.g. effect size) but are NOT recommended
- Hypotheses based responsiveness (construct responsiveness) and responsiveness based on a gold standard (criterion responsiveness) are good methods

Interpretation
- Minimal important change (MIC) and minimal important difference (MID) are different concepts according to COSMIN
- Anchor-based techniques uses an external anchor and are by far the most common and best methods
- MIC can be based on individuals or groups
- Mean value, difference in mean value and ROC-techniques are the most reported methods
- MIC_{IAPM} and MIC_{LCFA} are the most precise

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

33

Questions?

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS

SDU

34