Responsiveness and Minimal Important Change for core outcomes in low back pain – Statistical Analysis Plan

Date: 12.04.2021

Contributors:

Lars Christian Bråten, PhD

Monica Wigemyr, MSc

Anne Julsrud Haugen, PhD

Jan Sture Skouen, prof

Anne Froholdt, PhD

Maja Wilhelmsen, PhD

Ansgar Espeland, prof

John-Anker Zwart, prof

Kjersti Storheim, prof

Lars Grøvle, PhD

Jens Ivar Brox, prof

Raymond Ostelo, prof

Margreth Grotle, prof

This Statistical analysis plan is performed in accordance with the COSMIN checklist manual (http://fac.ksu.edu.sa/sites/default/files/cosmin_checklist_manual_v9.pdf)

Introduction:

Patient reported outcome measures (PROMs) are increasingly regarded essential in order to make research on low back pain (LBP) relevant for patients. Despite being used for several decades, there are still major difficulties in interpreting change scores of PROMs used in low back pain (1, 2). Which scores of the various PROMs that represent trivial, small but important, moderate or large change in outcome is still uncertain.

The Minimal Important Change (MIC) is defined as the smallest change in an outcome measure that patients perceive as important, and should be looked at separately for improvement or deterioration. We here use the term MIC referring to within-group change, in accordance with the COSMIN manual (3) (inconsistent use in the literature). Responsiveness refers to the ability of an instrument to detect change over time in the construct to be measured, and can be understood as the validity of the change score (3, 4). It is recommended to assess responsiveness and MIC in multiple settings, and on data from clinical trials (5). Further, the MIC and responsiveness of core PROMs in back pain could also differ when collected by electronic media (6), although there is evidence of little difference in metric qualities from paper versions (7). It is recommended to use anchor-based methods to assess both concepts (3, 5). For these analyses we will use the Global perceived effect as the anchor. There is a large variety in previous estimates of MIC values for Roland-Morris Disability Questionnaire (RMDQ, scores 0-24) (3 to 6), Oswestry Disability Index (ODI, scores 0-100) (13 to 20), LBP intensity Numerical Rating Scale (NRS, scores 0-10) (2 to 3) and EuroQol's health related quality of life (EQ5D, scores −0.59 to 1) (0.11 to 0.30), (2, 8, 9) probably due to a lack of consensus on methodology. Recently, a set of criteria to validate MIC estimates has been published (1), which we will use here in order to evaluate our results (criterion approach). Responsiveness is also suggested to be assessed with a construct approach, by defining a set of hypotheses, in accordance with COSMIN guidelines (10).

Domain	**Instrument**
Disability	• Roland-Morris Disability Questionnaire (RMDQ) • Oswestry Disability Index (ODI)
Pain intensity	• LBP intensity Numerical Rating Scale (NRS)
Quality of life	• EuroQol's health related quality of life (EQ5D)

The relevance (part of content validity) of RMDQ has been challenged (11). A qualitative study described patients with discordance between change in RMDQ score and a global perceived effect as an anchor (12). The study defined discordance as a different direction in change in RMDQ score over time compared to the global perceived effect. It found that this discordance could be due to patients referring to other specific domains than disability when answering the global perceived effect (12). The authors were concerned about using the Global perceived effect to calculate MIC scores (12). Further, global perceived effect might be more influenced by the current status than change in scores (13, 14). It is therefore important to quantify how common such discordance is and quantify whether discordance in one domain is related to discordance in other domains (disability, LBP intensity or quality of life). If there are many patients with discordance between the outcome domains and Global perceived effect, this could question the relevance of using Global perceived effect as anchor when calculating the MIC scores, in correlation analyses, and perhaps also the interpretation of the PROM itself.

Objectives:

Our overall aim is to evaluate the responsiveness and to estimate the minimal important change (MIC) of electronic-administered core PROMs for low back pain used in patients with chronic low back pain (LBP) and Modic changes. We used the following PROMs:

- The Norwegian version of the Roland Morris Disability Questionnaire (RMDQ), with a scale ranging from 0 to 24.(15)
- The Norwegian version of the Oswestry Disability Index (ODI) version 2.0 (scale range 0-100).(15, 16)
- LBP intensity (the mean of the following three 0–10 NRS scores; current LBP, worst LBP within the last 2 weeks, and usual/mean LBP within the last 2 weeks).(17)
- Health-related quality of life, EuroQoL-5D (EQ5D) version 2.0 (scale range -0.59 – 1.0).(18)
- Global perceived effect (7 point likert scale with alternatives: completely recovered, much improved, slightly improved, no change, slightly worsened, much worsened, and worse than ever)

Overview of the five specific objectives and corresponding analyses:

Objectives	Analyses
Assessment of assumptions	
1. Assess the percentage of patients with discordance between direction of change over time in scores of individual PROMs (RMDQ, ODI, LBP intensity and EQ5D) and the global perceived effect (assumption for B2)	A3
2. Assess the percentage of patients with discordance in direction of change over time scores in all PROMs (RMDQ, ODI, LBP intensity and EQ5D) relative to the Global perceived effect	A3 (+ A4)
Assessment of Responsiveness and MIC for the electronically administered outcomes RMDQ, ODI, Pain intensity and EQ5D	
3. Calculate Minimal important change (within group) for RMDQ, ODI, LBP intensity NRS and EQ5D	B2 (cut point) + A1
4. Assess responsiveness of RMDQ, ODI, LBP intensity NRS and EQ5D by construct approach	B1 Predefined hypotheses (table 6)
5. Compare responsiveness of RMDQ, ODI, LBP intensity and EQ5D by criterion approach against the global perceived effect	B2 (AUC)

RMDQ	Roland-Morris Disability Questionnaire
ODI	Oswestry Disability Index
NRS	Numerical Rating Scale
EQ5D	EuroQol’s health related quality of life

We intend to evaluate objective 2 in a separate paper.

Analyses:

All analyses will be performed on the whole cohort (amoxicillin + placebo group) of the AIM-study. We will analyze and report on follow-up data at 3 and 12 months for all PROMs and Global perceived effect. Patients with missing values in each PROM or Global perceived effect will be excluded in their respective analyses.

A. Assessment of assumptions

A1. Descriptive data
Box plot with descriptive data (median + percentiles 25 + 75) for each PROM on the following type of scores (presented on the y-axis):
-Absolute change; (baseline – 3 or 12 m value)
-Relative change; (baseline – 3 or 12 m value) / baseline value
Where the x-axis is all categories of the Global perceived effect.

A2. Correlations between absolute change in PROM score and the Global Perceived Effect
These analyses of correlations will be performed for each PROM and for both 3 and 12 months follow-up data.

A3. Pattern of discordance across all PROMs
Table with pattern of discordance in RMDQ, ODI, LBP intensity and EQ5D compared to the Global perceived effect (see Example table 1). There will be one table for each follow-up time of 3 and 12 months.
Discordance is defined as a difference between the direction of the change score (difference between score at 3 or 12 months and the score at baseline) and the direction of the Global perceived effect. That is, if Global perceived effect is scored as completely recovered/much improved/slightly improved, discordance is defined as a score of the PROM (at 3 and 12 months) equal or worse compared to the score at baseline. If Global perceived effect is scored as somewhat worse/much worsened/worse than ever, discordance is defined as a better score of the PROM at 3 or 12 months compared to the score at baseline.
Concordance is defined as not discordance. These analyses will only include patients with nonmissing values for RMDQ, ODI, LBP intensity, EQ5D and Global perceived effect (n=161 for follow-up time of 3 months and n=162 for follow-up time of 12 months).

Example table 1 (with invented numbers)

Pattern RMDQ, ODI, LBP intensity, EQ5D	N	%
0000		60
0001		4
0010		7
0011		1
0100		4
0101		2

0110		1
0111		1
1000		4
1001		3
1010		0
1011		0
1100		10
1101		0
1110		2
1111		1
Sum	161/162	100
Total discordance for each PROM		
RMDQ (1000+1001+1010+1011+1100+1101+1110+1111)		20
ODI (0100+0101+0110+0111+1100+1101+1110+1111)		31
LBP intensity (0010+0011+0110+0111+1010+1011+1110+1111)		13
EQ5D (0001+0011+0101+0111+1001+1011+1101+1111)		12

Pattern describes whether there is discordance (1) or concordance (0) of the PROMs compared to the Global perceived effect for the following order of PROMs: RMDQ, ODI, LBP intensity, EQ5D. Discordance is defined as a difference between the score at 3 or 12 months and the score at baseline in the opposite direction as suggested by the Global perceived effect.

A4. Ability of individual and combination of PROMs to separate better vs unchanged/worse

Table with individual and combination of PROMs and the Global perceived effect, both dichotomized into better and unchanged/worse (see Example table 2).

Example table 2 (with invented numbers for RMDQ)

		Global perceived effect Better	Global perceived effect Unchanged/ Worse	P-value	sensitivity	specificity	AUC ROC With CI
RMDQ	N=167	-	-				
	Better	90	17	-	-	-	-
	Unchanged/ Worse	20	40	-	-	-	-
ODI	N=						
	Better						
	Unchanged/ Worse						
LBP intensity	N=						
	Better						
	Unchanged/ Worse						
EQ5D	N=						

	Better						
	Unchanged/ Worse						
RMDQ + LBP intensity	N=						
	Better RMDQ or pain int						
	Worse RMDQ and pain int						
ODI + LBP intensity	N=						
	Better ODI or pain int						
	Worse ODI and pain int						

B. Assessment of individual PROMs and types of scores

The following analyses (B1-B3) will be performed on each of the PROMs RMDQ, ODI, LBP intensity and EQ5D:

B1. Analyses of predefined hypotheses to test responsiveness of PROMs

See table 6 (all refers to 12 months follow-up).

B2. ROC curve analysers

Global perceived effect will be dichotomized into:
a. completely recovered/much improved/slightly improved vs no change (n=143) (worse categories (n=26) excluded)
b. completely recovered/much improved vs slightly improved/no change (n=143) (worse categories (n=26) excluded)

Worse categories will be excluded as we focus on MIC and responsiveness for improvement only (not deterioration) and consider including the worse categories would introduce unnecessary noise (19). For **a** and **b**, we will present a table with columns: N, cut-point (the Nearest method, recommended by NIH (20)), sensitivity, specificity, area under the ROC curve (AUC ROC)

And for each follow-up of 3 and 12 months, rows (types of score):

1. Absolute change; (baseline – follow-up value)
2. Relative change; (baseline – follow-up value) / baseline value

We will report the cut-point on absolute and relative (in percent) change on **analysis a** and **b** at 3 and 12 months follow-up for objective 1 (MIC), and use the AUC to compare PROMs at 12 months follow-up for objective 5 (responsiveness) (table 3). We will estimate 95% confidence interval for cut-point and AUC by bootstrapping.

B3. Logistic regression to estimate MIC

We will as a sensitivity analysis for estimating the MIC (ROC analyses a and b in B2) do logistic regression and report the change value associated with a likelihood ratio of 1 (21). In these sensitivity analyses we will use the dichotomized global perceived effect values (a and b, respectively) as dependent variable and absolute change from baseline to 12 months follow-up as independent variable. We will also perform an adjusted logistic regression model in order to check the assumption that baseline values and proportion of improved patients are not unduly influencing the MIC estimates (22). In this adjusted model, we will add the baseline value dichotomized into high-baseline (larger than median baseline value) and low-baseline (smaller or equal to median baseline value), and the interaction term absolute change x baseline value (dichotomized). We will further adjust for the proportion of improved patients. The following formula to adjust for the proportion of improved patients will be used (22):

$$MICa = MICp\ - (0.090 + 0.103 \times C) \times SDc \times \ln(imp)$$

Where MICa = adjusted minimal important change; MICp = minimal important change predicted by the logistic regression model; C = point-biserial correlation between the PROM change score and the anchor; SDc = standard deviation of the PROM change score; ln(imp) = natural logarithm of [proportion improved/(1 − proportion improved)]

We will present separate results for high-baseline and low-baseline groups from the adjusted logistic regression model.

How the analyses will answer the objectives

Assumptions

For objective 3 to 5, we will require a correlation between absolute change in PROM score and the Global Perceived Effect (analyses A2), for each follow-up of 3 and 12 months, of at least 0.5{Guyatt, 2002 #2773}. If this requirement is not met for any timepoint, we will not report analyses for the relevant timepoint for objective 3 to 5.

For objective 1, we will assess the percentage of patients with discordance between direction of change over time in scores for individual PROMs and the global perceived effect by calculating

- The total number and percentage of patients with discordance for each PROM (Total discordance for each PROM in example table 1 in A3).

We will compare this total discordance across the three domains, for both 3 and 12 months follow-up. This assessment is meant to answer how much each of the three main domains (disability, LBP intensity or quality of life) contributes to patients' reporting of the Global perceived effect. We will consider high discordance for each PROM as a possible threat to the underlying assumption in the B2 analyses. However, we intend to report the B2 results regardless of results of these analyses of discordance, as these are to our knowledge new and not universally accepted as assumptions.

For objective 2, we will assess the percentage of patients with discordance in all PROMs relative to the Global perceived effect by calculating

- The number and percentage of patients with patterns number 0111, 1011 and 1111 (in example table 1) for both 3 and 12 months follow-up.

A large number of patients with such discordance in all PROMs/domains could suggest weaknesses with Global perceived effect as an anchor (eg. patients could be unable to remember accurately 12 months back). Alternatively, it could suggest we are not able to pick up improvement regarded relevant by patients when we measure disability, LBP intensity and quality of life.

In addition, we will assess how the discordance in direction of change over time scores for disability outcomes (RMDQ and ODI) vs Global perceived effect are associated with the same discordance for LBP intensity by calculating

- The number and percentage of patients with discordance in only one of RMDQ and LBP intensity (Sum of 0010+0011+0110+ 0111+1000+1001+1100+1101 in example table 1).
- The number and percentage of patients with discordance in only one of ODI and LBP intensity (Sum of 0010+0011+0100+ 0101+1010+1011+1100+1101 in example table 1).

We will also assess the discriminative ability of all individual PROMs and the combination of RMDQ/ODI with NRS, to separate better from unchanged/worse in the Global perceived effect, by calculating AUC with confidence intervals using bootstrapping (A4).

For objective 3 (calculation of MIC), for those conditions in which the requirements mentioned in assumptions above are met, we will analyse:

- For each PROM, for 3 and 12 months follow-up, and for absolute and relative change, the MIC calculated using the cut-off score (**analysis a** and **b** in B2)(recommended by NIH (20)). We will view analysis **a** as of general interest for common back pain interventions, and analysis **b** as of particular interest for interventions with higher risk of adverse reactions (eg surgery or antibiotic treatment). The credibility of these findings will be evaluated based on 5 criteria (table 4 and 5), using the credibility criteria provided in this paper (1). However, criteria 3 (item 3 in table 4) will be viewed as a prerequisite for the MIC calculations (see assumptions above). We will compare results from the analyses of 3 and 12 months follow-up, to assess if the amount of elapsed time between baseline and follow-up measurement influence the estimates (see item 6 in table 4). As sensitivity analyses, we will do:
    - Logistic regression (B3), to assess assumptions of no unduly influence of baseline values or proportion of improved patients.
- For each PROM, mean and 25th percentile (p25) in patients who report 1.slightly improved and 2. much improved in the Global perceived effect (table 3). Equivalent to analyses **a** and **b** above, we will view analysis **1** as of general interest for common back pain interventions, and analysis **2** as of particular interest for interventions with higher risk of adverse reactions (eg surgery or antibiotic treatment).

For objective 4 (responsiveness assessed by construct approach), we will assess responsiveness by evaluating 10 predefined hypotheses on correlation of absolute change scores, as recommended by the COSMIN panel (3). We will assess the hypothesis by the calculated correlations, standardized mean differences (SMD) and standardized response means (SRM), and no p-values will be reported. We will require 75% of the hypotheses for each outcome (eg. 5 out of 6) confirmed in order to state good responsiveness (10).

For objective 5 (responsiveness assessed by criterion approach), we will compare responsiveness of RMDQ, ODI, Pain intensity and EQ5D by criterion approach against the global perceived effect, for both absolute and relative change by comparing AUC values in analyses **a** in B2. We will consider AUC>0.7 as adequate (23, 24).

Table 3. Summary table of analyses for each objective

Objective	Main analyses
Assessment of assumptions	
1. Assess the percentage of patients with discordance between direction of change over time in scores of individual PROMs (RMDQ, ODI, LBP intensity and EQ5D) and the global perceived effect (assumption for B2)	• Total discordance for each PROM for Better and Unchanged/worse at 3 and 12 months follow-up (A3).
2. Assess the percentage of patients with discordance in direction of change over time scores in all PROMs (RMDQ, ODI, LBP intensity and EQ5D) relative to the Global perceived effect	• The number and percentage of patients with patterns number 0111, 1011 and 1111 (in example table 1) for both 3 and 12 months follow-up (A3).
Assessment of Responsiveness and MIC for the electronically administered outcomes RMDQ, ODI, Pain intensity and EQ5D	
3. Calculate Minimal important change (within group) for RMDQ, ODI, LBP intensity NRS and EQ5D	• Cut-off scores for absolute and relative change (ROC **analysis a and b** in B2) at 3 and 12 months follow-up • Median + p25 (A1) for the 1.slightly improved group and 2.much improved group, at 3 and 12 months follow-up • Sensitivity analysis: Logistic regression (B3), at 3 and 12 months follow-up
4. Assess responsiveness of RMDQ, ODI, LBP intensity NRS and EQ5D by construct approach	• 6 hypothesis for each outcome (B1 Table 6)
5. Compare responsiveness of RMDQ, ODI, LBP intensity and EQ5D by criterion approach against the global perceived effect	• AUC scores for absolute and relative change (**analysis a** in B2), at 12 months follow-up

Table 4. Assessing credibility of the estimated Minimal Important Change (based on (1)):

Signalling question	Response options	
	High credibility	Low credibility
Core Criteria	Assessment	
Item 1: Is the patient or necessary proxy responding directly to both the patient reported outcome measure and the anchor?	Yes	No/impossible to tell
Item 2: Is the anchor easily understandable and relevant for patients or necessary proxy?	Definitely yes/to a great extent	Definitely no/not so much/impossible to tell
Item 3: Has the anchor shown good correlation with the patient reported outcome measure?		
Item 4: Is the MIC precise?		
Item 5: Does the threshold or difference between groups on the anchor used to estimate the MIC reflect a small but important difference?		
Additional criteria for transition rating anchors		
Item 6: Is the amount of elapsed time between baseline and follow-up measurement for MIC (termed MID in original list, but meaning is the same) estimation optimal?	Definitely yes/to a great extent	Definitely no/not so much/impossible to tell
Item 7: Does the transition item have a satisfactory correlation with the PROM score at follow-up?		
Item 8: Does the transition item correlate with the PROM score at baseline?		
Item 9: Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PROM score at follow-up?		

Table 5. Considerations for judging whether the minimal important difference represents a small but important difference (item 5 in table 4 (1))

1. What is the original scale of the anchor and is it transformed in any way?
2. Does the scale (or transformed scale) of the anchor capture variability in the underlying construct?
3. What is the threshold used or comparison being made on the anchor? Does this threshold or comparison represent a difference that is minimally important?

4. Does the analytical method ensure that the minimal important difference represents a small but important difference?

Table 6. Predefined hypotheses to test responsiveness of PROMs

Hypothesis	**Exp value**	**Calc value (with CI)**
RMDQ		
The correlation between absolute change in RMDQ score and the absolute change in ODI score at 12 months is at least strong and positive as they measure the same construct(25, 26)	r ≥0.7	
The correlation between absolute change in RMDQ score and the absolute change in LBP intensity NRS score at 12 months is moderate and positive(25, 26)	r ≥0.3 and <0.7	
The correlation between absolute change in RMDQ score and the absolute change in EQ5D score at 12 months is moderate and positive (27)	r ≥0.3 and <0.7	
The standardized response mean in absolute RMDQ change score is less than 0.2 for those who scored no change on the Global Perceived Effect at 12 months (25)	SRM <0.2	
The standardized response mean in absolute RMDQ change score is more than 0.2 for those who scored slightly improved on the Global Perceived Effect at 12 months (25)	SRM >0.2	
The standardized response mean in absolute RMDQ change score is more than 0.5 for those who scored much improved on the Global Perceived Effect at 12 months (25)	SRM >0.5	
ODI		
The correlation between absolute change in ODI score and the absolute change in RMDQ score at 12 months is at least strong and positive as they measure the same construct(25, 26)	r ≥0.7	
The correlation between absolute change in ODI score and the absolute change in LBP intensity NRS score at 12 months is moderate and positive(25, 26)	r ≥0.3 and <0.7	
The correlation between absolute change in ODI score and the absolute change in EQ5D score at 12 months is moderate and positive (27)	r ≥0.3 and <0.7	
The standardized response mean in absolute ODI change score is less than 0.2 for those who scored no change on the Global Perceived Effect at 12 months (8, 25)	SRM <0.2	
The standardized response mean in absolute ODI change score is more than 0.2 for those who scored slightly improved on the Global Perceived Effect at 12 months (8, 25)	SRM >0.2	

The standardized response mean in absolute ODI change score is more than 0.5 for those who scored much improved on the Global Perceived Effect at 12 months (8, 25)	SRM >0.5	
Pain intensity (NRS)		
The correlation between absolute change in LBP intensity NRS score and the absolute change in ODI score at 12 months is moderate and positive (25, 26)	r ≥0.3 and <0.7	
The correlation between absolute change in LBP intensity NRS score and the absolute change in RMDQ score at 12 months is moderate and positive(25, 26)	r ≥0.3 and <0.7	
The correlation between absolute change in LBP intensity NRS score and the absolute change in EQ5D score at 12 months is moderate and positive (27)	r ≥0.3 and <0.7	
The standardized response mean in absolute LBP intensity NRS change score is less than 0.2 for those who scored no change on the Global Perceived Effect at 12 months (25)	SRM <0.2	
The standardized response mean in absolute LBP intensity NRS change score is more than 0.2 for those who scored slightly improved on the Global Perceived Effect at 12 months (8, 25)	SRM >0.2	
The standardized response mean in absolute LBP intensity NRS change score is more than 0.5 for those who scored much improved on the Global Perceived Effect at 12 months (8, 25)	SRM >0.5	
EQ5D		
The correlation between absolute change in EQ5D score and the absolute change in ODI score at 12 months is moderate and positive (25, 26)	r ≥0.3 and <0.7	
The correlation between absolute change in EQ5D score and the absolute change in LBP intensity NRS score at 12 months is moderate and positive(25-27)	r ≥0.3 and <0.7	
The correlation between absolute change in EQ5D score and the absolute change in RMDQ score at 12 months is moderate and positive (27)	r ≥0.3 and <0.7	
The standardized response mean in absolute EQ5D change score is less than 0.2 for those who scored no change on the Global Perceived Effect at 12 months (8)	SRM <0.2	
The standardized response mean in absolute EQ5D change score is more than 0.2 for those who scored slightly improved on the Global Perceived Effect at 12 months (8)	SRM >0.2	
The standardized response mean in absolute EQ5D change score is more than 0.5 for those who scored much improved on the Global Perceived Effect at 12 months (8)	SRM >0.5	

SRM (Standardized response mean) - the average difference divided by the standard deviation of the differences between the paired measurements

References:

1. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. bmj. 2020;369.
2. Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting Change Scores for Pain and Functional Status in Low Back Pain: Towards International Consensus Regarding Minimal Important Change. Spine. 2008;33(1):90-4.
3. Mokkink LB TC, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. COSMIN checklist manual COSMIN panel; 2012 [Available from: http://fac.ksu.edu.sa/sites/default/files/cosmin_checklist_manual_v9.pdf.
4. Terwee C, Dekker F, Wiersinga W, Prummel M, Bossuyt P. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Quality of life research. 2003;12(4):349-62.
5. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of clinical epidemiology. 2008;61(2):102-9.
6. Henschke N, van Enst A, Froud R, Ostelo RW. Responder analyses in randomised controlled trials for chronic low back pain: an overview of currently used methods. European Spine Journal. 2014;23(4):772-8.
7. Froud R, Fawkes C, Foss J, Underwood M, Carnes D. Responsiveness, reliability, and minimally important and minimal detectable changes of 3 electronic patient-reported outcome measures for low back pain: validation study. Journal of medical Internet research. 2018;20(10):e272.
8. Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JI, Hermansen E, et al. Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis. BMC Musculoskeletal Disorders. 2019;20(1):31.
9. Solberg T, Johnsen LG, Nygaard ØP, Grotle M. Can we define success criteria for lumbar disc surgery? Estimates for a substantial amount of improvement in core outcome measures. Acta orthopaedica. 2013;84(2):196-201.
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology. 2010;63(7):737-45.
11. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. Journal of Clinical Epidemiology. 2018;95:73-93.
12. Froud R, Ellard D, Patel S, Eldridge S, Underwood M. Primary outcome measure use in back pain trials may need radical reassessment. BMC musculoskeletal disorders. 2015;16:88.
13. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. Journal of clinical epidemiology. 2010;63(7):760-6. e1.
14. Grøvle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. Journal of clinical epidemiology. 2014;67(5):508-15.
15. Grotle M, Brox J, Vollestad N. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. Journal of rehabilitation medicine. 2003;35(5):241-7.
16. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine. 2000;25(22):2940.

17. Dworkin HR, Turk CD, Farrar TJ, Haythornthwaite AJ, Jensen PM, Katz PN, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(12):9-19.
18. EuroQol - a new facility for the measurement of health-related quality of life. Health policy. 1990;16(3):199-208.
19. De Vet HC, Ostelo RW, Terwee CB, Van Der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Quality of life research. 2007;16(1):131.
20. Dutmer AL, Reneman MF, Preuper HRS, Wolff AP, Speijer BL, Soer R. The NIH minimal dataset for chronic low back pain: responsiveness and minimal clinically important change. Spine. 2019;44(20):E1211.
21. Terluin B, Eekhout I, Terwee CB, de Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. Journal of Clinical Epidemiology. 2015;68(12):1388-96.
22. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. Journal of Clinical Epidemiology. 2017;83:90-100.
23. de Vet H TC, Mokkink L, Knol D. . Measurement in Medicine Cambridge: Cambridge University Press; 2011.
24. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42.
25. Grotle M, Brox JI, Vøllestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine. 2004;29(21):E492-E501.
26. Yao M, Xu B-p, Li Z-j, Zhu S, Tian Z-r, Li D-h, et al. A comparison between the low back pain scales for patients with lumbar disc herniation: validity, reliability, and responsiveness. Health and Quality of Life Outcomes. 2020;18(1):1-12.
27. Soer R, Reneman MF, Speijer BL, Coppes MH, Vroomen PC. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. The Spine Journal. 2012;12(11):1035-9.