Journal of Clinical Epidemiology 151 (2022) 75−87

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Minimal important change was on the lower spectrum of previous estimates and responsiveness was sufficient for core outcomes in chronic low back pain

Lars Christian Haugli Bråten[a,*], Lars Grøvle[b], Monica Wigemyr[a], Maja Wilhelmsen[c,d], Elisabeth Gjefsen[a,e], Ansgar Espeland[f,g], Anne Julsrud Haugen[b], Jan Sture Skouen[h], Jens Ivar Brox[i], John-Anker Zwart[a,e], Kjersti Storheim[a,l], Raymond WJG. Ostelo[j,k,l], Margreth Grotle[a,l]

[a]*Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital HF, Ulleval, Bygg 37b, Postbox 4956, Nydalen, 0424, Oslo, Norway*

[b]*Department of Rheumatology, Østfold Hospital Trust, PB 300, 1714, Grålum, Norway*

[c]*Department of Rehabilitation, University Hospital of North Norway, P.O. Box 100, 9038 Tromsø, Norway*

[d]*Faculty of Health Sciences, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway*

[e]*Faculty of Medicine, University of Oslo, P.O. Box 1072 Blindern, 0316, Oslo, Norway*

[f]*Department of Radiology, Haukeland University Hospital, Jonas Liesvei 65, 5021 Bergen, Norway*

[g]*Department of Clinical Medicine, University of Bergen, P.O. Box 7804, 5020, Bergen, Norway*

[h]*Department of Physical Medicine and Rehabilitation, Haukeland University Hospital, Helse Bergen HF, Box 1, 5021 Bergen, Norway*

[i]*Department of Physical Medicine and Rehabilitation, Oslo University Hospital HF, Ulleval, Postbox 4956, Nydalen, 0424, Oslo, Norway*

[j]*Department of Health Sciences, Faculty of Science, VU University Amsterdam, Amsterdam Movement Sciences Research Institute Amsterdam, Amsterdam, Netherlands*

[k]*Department of Epidemiology and Data Science, Amsterdam University Medical Centre, Location VUmc, Amsterdam, Netherlands*

[l]*Oslo Metropolitan University, Department of Physiotherapy, PO box 4 St. Olavs plass, NO-0130 Oslo, Norway*

Accepted 21 July 2022; Published online 1 August 2022

Abstract

Objectives: The objective of this study was to estimate the minimal important change (MIC) and responsiveness of core patient reported outcome measures for chronic low back pain (LBP) and Modic changes.

Study Design and Setting: In the Antibiotics in Modic changes (AIM) trial we measured disability (RMDQ, ODI), LBP intensity (NRS) and health-related quality of life (EQ5D) electronically at baseline, three- and 12-month follow-up. MICs were estimated using Receiver Operating Curve (ROC) curve and Predictive modeling analyses against the global perceived effect. Credibility of the estimates

The AIM-study group (see list of authors under acknowledgment).

Author contributions: Lars Christian Haugli Bråten: Conceptualization, Formal analysis, Writing - original draft Lars Grøvle: Investigation, Methodology, Writing - review & editing Monica Wigemyr: Data Curation, Resources, Writing - review & editing Maja Wilhelmsen: Investigation, Writing - review & editing Elisabeth Gjefsen: Writing - review & editing, Visualization Ansgar Espeland: Methodology, Writing - review & editing Anne Julsrud Haugen: Investigation, Writing - review & editing Jan Sture Skouen: Investigation, Writing - review & editing Jens Ivar Brox: Methodology, Supervision, Writing - review & editing John-Anker Zwart: Project administration, Funding acquisition, Writing - review & editing Kjersti Storheim: Project administration, Funding acquisition, Writing - review & editing Raymond WJG Ostelo: Supervision, Methodology, Conceptualization, Writing - review & editing Margreth Grotle: Supervision, Methodology, Conceptualization, Writing - review & editing.

Conflict of interest: All authors declare no conflict of interest.

Data sharing statement: Requests to access data should be addressed to kjersti.storheim@medisin.uio.no. De-identified individual participant data (including data dictionary) will be available to medical researchers by request in accordance with local registration and ethical approval, until 1st of July, 2029. All proposals requesting data access will need to specify an analysis plan and will need approval of the scientific board before any data can be released.

Declaration of interests: The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

* Corresponding author: Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital HF, Ulleval, Bygg 37b, Postbox 4956, Nydalen, 0424, Norway, Tel.: +47 22117769; fax: +47 23016150.

E-mail address: l.c.h.braten@studmed.uio.no (L.C.H. Bråten).

https://doi.org/10.1016/j.jclinepi.2022.07.012

was assessed by a standardized set of criteria. Responsiveness was assessed by a construct and criterion approach according to COSMIN guidelines.

Results: The MIC estimates of RMDQ, ODI and NRS scores varied between a 15−40% reduction, depending on including ''slightly improved'' in the definition of MIC or not. The MIC estimates for EQ5D were lower. The credibility of the estimates was moderate. For responsiveness, five out of six hypotheses were confirmed and AUC was >0.7 for all PROMs.

Conclusion: When evaluated in a clinical trial of patients with chronic LBP and Modic changes, MIC thresholds for all PROMs were on the lower spectrum of previous estimates, varying depending on the definition of MIC. Responsiveness was sufficient.

Keywords: Minimal important change; Low back pain; Responsiveness; Patient reported outcome measure; Roland-Morris disability questannaire; Oswestry disability index; Pain intensity numerical rating scale; EuroQol's health related quality of life; Predictive modeling

1. Introduction

Patient reported outcome measures (PROMs) are essential in low back pain (LBP) research, contribute to results being relevant for patients and aid in informing shared decision-making. Estimates of Minimal Important Change (MIC), defined as the smallest change in an outcome measure that patients perceive as important, are central for interpreting results from prospective studies and trials. 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 [1,2].

Despite being used for several decades, there are still major difficulties in interpreting change scores of PROMs used in LBP research [3,4]. There is a large variety in estimates of MIC values for recommended core outcomes [5]. MIC estimates of the Roland-Morris Disability Questionnaire (RMDQ, scores 0-24) vary from 3 to 6, the Oswestry Disability Index (ODI, scores 0-100) from 13 to 20, the LBP intensity Numerical Rating Scale (NRS, scores 0-10) from 2 to 3, and for the EuroQol's health related quality of life (EQ5D, scores −0.59 to 1) from 0.11 to 0.30 [4,6,7]. For MIC thresholds reported as relative change, a 30% improvement is often recommended in many PROMs [4,8]. The large variety may be explained by different methods and populations used to estimate the MIC, or possibly biased or misleading results. In order to guide an evaluation of whether estimates might be misleading, a standardized set of credibility criteria has recently been developed [3]. There are inconsistent results on responsiveness of core PROMs in back pain in previous studies [9,10].

The objective of this study was to estimate MICs across methods and responsiveness of core PROMs in patients with chronic LBP and Modic changes.

2. Materials and methods

2.1. Setting

The present study is based on data from the AIM-trial (Antibiotics in Modic changes) performed in six hospital outpatient clinics in Norway between June 2015 and September 2017 (11). Patients with age 18−65 years, chronic LBP for more than six months, a minimum pain intensity of at least five on 0−10 numerical rating scales (current LBP, worst LBP within the last 2 weeks, and usual/mean LBP within the last two weeks), a lumbar disc herniation in the preceding two years, and type 1 or 2 vertebral bone marrow changes (Modic changes) on magnetic resonance imaging (MRI), were eligible for inclusion. The full list of eligibility criteria and trial methods were published elsewhere [11]. In total 180 patients were randomized to three months treatment of either oral amoxicillin or placebo, and then followed-up for another nine months. All patients provided written informed consent and the trial was approved by Regional Committees for Medical Research Ethics South East Norway (2014/158/REK sør-øst). Funding was granted by governmental organizations (Helse Sør-Øst and Helse Vest), which had no part in the planning, performing, or reporting of the trial.

2.2. PROMS and external anchor

We collected the following PROMs by using a web-based data capture system (Viedoc): The Norwegian version of the RMDQ, a reflective model (all items a manifestation of the same construct) with a scale ranging from 0 to 24 (higher scores indicate more pain-related disability) [12,13]; the Norwegian version of the ODI version 2.0, a reflective model, scale range 0-100 (higher scores indicate more pain-related disability) [12,14]; LBP intensity by 0-10 NRS (the mean of current LBP, worst, and usual/mean LBP within the last two weeks, higher scores indicate more pain), a reflective model [15]; and health-related quality of life, EuroQoL-5D-5 L (EQ5D) version 2.0, a formative model (the items together form the construct), with scale range −0.59 to 1.0 (the maximum score of one indicates the best health-related quality of life) [16].

The external anchor for assessing responsiveness and MIC estimates was the Global perceived effect (GPE) scale. The patients were asked ''Compared to before treatment, how is your back pain now?''. A 7-point Likert scale

What is new?

Key findings

- In patients with chronic LBP (with Modic changes) the MIC thresholds of RMDQ, ODI and LBP intensity scores varied between 15% and 40% reduction, depending on definition of MIC (including those who reported ''slightly improved'' or not). The MIC estimates had moderate credibility. Responsiveness was sufficient for all these outcome measures.

What this adds to what was known?

- MIC thresholds of RMDQ, ODI and LBP intensity scores are on the lower spectrum of previous estimates in patients with chronic LBP and Modic changes in an outpatient hospital setting.
- The present study is the first to assess credibility of MIC estimates in LBP by a standardized set of criteria.

What is the implication and what should change now?

- Our estimated within-group MIC thresholds could be used to prespecify a definition of a responder in responder analyses, or with caution to assess individual changes in the clinical setting.

was based on the following options: 1.completely recovered; 2.much improved; 3.slightly improved; 4.no change; 5.slightly worsened; 6.much worsened; 7.worse than ever.

The PROMs were measured at baseline, three- and 12-month follow-up. GPE was measured at three- and 12-month. Patients received text message reminders to respond to the PROMs. In a few cases patients used a paper version.

2.3. Analyses

This paper is reported in accordance with the COSMIN guidelines for studies on measurement properties of PROMs [17]. The AIM trial and the statistical analysis plan for the present study were registered at ClinicalTrials.gov (NCT02323412). Analyses not predefined, are marked as such. Analyses were performed on the whole AIM cohort (amoxicillin + placebo group) using Stata 16. In patients with less than 30% (<8 of the 24) missing items on the RMDQ, a score was calculated by using the mean values of nonmissing items. For ODI there is a similar mechanism. Other missing values after mean imputation or in LBP intensity or EQ5D were excluded.

2.3.1. Minimal important change

We stress that the MIC estimates refer to within-group changes and should not be confused with estimates of between-group differences. A common and widely accepted method to estimate MIC thresholds is by using an external anchor [3,18]. Analyses were performed for three- and 12-month changes as follows:

1. The primary analyses were based on Receiver Operating Curve (ROC) curve analyses. We analyzed cut-points (using 'nearest' [19] in Stata), sensitivity, specificity and AUC for absolute change (change between baseline and follow-up score with improvement of symptoms given positive values) and relative change (change from baseline to follow up as percentage of baseline score with improvement of symptoms given positive values).
2. As secondary analyses we calculated medians +25th percentile (p25) of change scores for the slightly improved and the much improved groups on the GPE.
3. Sensitivity analyses: To check the assumption that baseline values and the proportions of improved patients not unduly influenced the MIC estimates [20] we performed predictive modeling using logistic regression as described by Terluin et al. [21]. These analyses are recommended as they provide more precise estimates than ROC curve analyses [17]. We further performed predictive modeling using random-effects logistic regression (not predefined), otherwise exactly as described by Terluin et al. [21], that combine three- and 12-month data to further narrow down the confidence intervals.

As a prerequisite for the analyses we checked that the correlation between the absolute change in PROM scores and the GPE was at least 0.5 [22].

We investigated two frequently used definitions for a minimal important change in the external anchor (the global perceived effect), with and without the ''slightly improved'' category. In analyses 1 and 3 the GPE was dichotomized into either 'a' —completely recovered/much improved/slightly improved vs no change (worse categories excluded) or 'b' —completely recovered/much improved vs slightly improved/no change (worse categories excluded). We consider 'a' as relevant for low-risk interventions and 'b' as relevant for interventions with higher risk of adverse reactions (eg., surgery or antibiotic treatment). For cut-point scores and logistic regression estimates we also performed separate analyses in patients with baseline values lower than the median and in patients with values higher than median. Confidence intervals for cut-point scores and logistic regression estimates were calculated by bootstrap replications.

The credibility of our findings was evaluated by a standardized set of criteria that resulted in an overall judgment (low/moderate/high credibility) as suggested [3]. Further details of our analyses are described in the Statistical analysis plan [23].

Table 1. Patient reported outcomes at baseline, 3 months and 1 year, with information on missingness

		Total cohort including all categories of GPE				Excluding worse categories of GPE			
PROM	Follow-up	Completed PROM (*n*)	Mean (SD)	Completed PROM and GPE	Missing PROM and/or GPE (*n*, %)	Completed PROM (*n*)	Mean (SD)	Completed PROM and GPE	Missing PROM and/or GPE (*n*, %)
RMDQ (0–24)	Baseline	178	12.8 (4.2)						
RMDQ (0–24)	3 months	172	11.4 (5.3)	166	14 (8)	156	10.9 (5.2)	154	26 (14)
RMDQ (0–24)	1 year	169	9.9 (5.9)	167	13 (7)	143	8.8 (5.6)	141	39 (22)
ODI (0–100)	Baseline	177	31.8 (10.8)						
ODI (0–100)	3 months	171	28.5 (13.0)	167	13 (7)	157	27.2 (12.5)	155	25 (14)
ODI (0–100)	1 year	169	26.6 (14.6)	166	14 (8)	143	23.8 (13.3)	141	39 (22)
LBP intensity (0–10)	Baseline	178	6.3 (1.4)						
LBP intensity (0–10)	3 months	170	5.3 (2.1)	166	14 (8)	156	5.1 (2.1)	154	26 (14)
LBP intensity (0–10)	1 year	169	5.0 (2.3)	167	13 (7)	143	4.5 (2.2)	141	39 (22)
EQ5D (−0.59 to 1.0)	Baseline	180	0.5 (0.2)						
EQ5D (−0.59 to 1.0)	3 months	168	0.6 (0.2)	167	13 (7)	155	0.6 (0.2)	155	25 (14)
EQ5D (−0.59 to 1.0)	1 year	167	0.6 (0.2)	167	13 (7)	142	0.6 (0.2)	142	38 (21)

RMDQ, Roland Morris Disability Questionnaire; ODI, Oswestry Disability Index; LBP (low back pain) intensity, measured by numerical rating scale; EQ5D, Health-related quality of life, EuroQoL-5D; GPE, Global perceived effect.

2.3.2. *Responsiveness*

We assessed the responsiveness by two approaches:

1. Construct approach: Six hypotheses regarding correlation of the 12 months absolute changes of each PROM with other PROMS and expected standardized response means within categories of the GPE were assessed (see Table 3) [1]. Standardized response mean is the average change divided by the standard deviation of the change. To state sufficient responsiveness we required 75% (5 out of 6) of the hypotheses for each PROM confirmed [24].
2. Criterion approach: Area under the curve (AUC) for absolute and relative change scores using GPE at 12 months follow-up were assessed, dichotomized as for a above, as reference variable (We required an AUC$>$0.7 to state sufficient responsiveness [24,25].

We calculated power for the responsiveness analysis only, as standard errors for MIC estimates are not readily calculable. Based on the subjects who completed all PROMs at the 12-month follow-up, of whom 77 reported completely recovered/much improved/slightly improved and 61 reported no change, and assuming an AUC of 80%, we would be able to estimate an AUC with a 95% confidence interval of ±7.3% [26].

3. Results

This sample included 180 patients with median (IQR) pain duration three (1.5-6.3) years and a mean (SD) 45 [9] years of age, of whom 105 (58%) were women. At three months follow-up we excluded 12 patients who were worse and 11 patients with missing values in GPE. Out of the remaining 157 patients, 66 (42%) and 31 (20%) improved, defined respectively, as including and excluding the slightly improved category of GPE. At 12 months follow-up 26 patients were worse, and 80 (56%) and 42 (29%) out of 143 patients improved, defined respectively, as including and excluding the slightly improved category of GPE. The distribution of absolute and relative change scores within each category of GPE are shown for each PROM in Figure 1. The correlations (absolute values) between change in each PROMs score and the GPE was $>$0.5 (Table S1). PROM scores at baseline, three and 12 months, and information on missingness are reported in Table 1.

3.1. *Estimates of minimal important change (MIC)*

Figure 2 presents the MIC estimates for absolute and relative change scores of the PROMs across the methods used (ROC and Predictive modeling analyses) for the two time points (3- and 12-month follow-up) as well as for the two cut-offs on the GPE (analyses ‘a’ and ‘b’). The ROC cut-point scores for analyses ‘a’ on absolute change at 3 months follow-up was 1.5 (95% CI 0.0 to 3.0) for RMDQ, 3.0 (95% CI 1.2 to 4.8) for ODI, 1.3 (95% CI 0.8 to 1.8) for LBP intensity and 0.03 (95% CI 0.01 to 0.05) for EQ-5D. Generally, the MIC estimates were larger at 12 months than at 3 months follow-up, and for analyses ‘b’ vs. ‘a’. When slightly improved was included in the definition of improved (analyses ‘a’) the MIC estimates of relative changes were 15-25% compared to 25-40% when slightly improved was not included (analyses ‘b’). Relative changes on the EQ-5D were not calculated as infinite values occur when baseline values are zero.

Table 2. Assessing credibility of the estimated Minimal Important Change [3]

	Credibility assessment with rationale	
Signaling questions	**3-month follow-up**	**12-month follow-up**
Core Criteria		
Item 1: Is the patient or necessary proxy responding directly to both the patient reported outcome measure and the anchor?	Yes, patients completed all questionnaires.	Yes, patients completed all questionnaires.
Item 2: Is the anchor easily understandable and relevant for patients or necessary proxy?	Definitely yes. The anchor is a transition rating that asks, ''Compared to before starting treatment, how is your back pain now?'' The seven responses were ''completely recovered'', ''much improved'', ''slightly improved'', ''no change'', ''slightly worsened'', ''much worsened'', and ''worse than ever''.	Definitely yes. The anchor is a transition rating that asks, ''Compared to before starting treatment, how is your back pain now?'' The seven responses were ''completely recovered'', ''much improved'', ''slightly improved'', ''no change'', ''slightly worsened'', ''much worsened'', and ''worse than ever''.
Item 3: Has the anchor shown good correlation with the [change in] patient reported outcome measure?	Definitely yes for LBP intensity NRS (0.73). To a great extent for RMDQ (0.59), ODI (0.60) and EQ-5D[a] (−0.57).	Definitely yes for LBP intensity NRS (0.76) RMDQ (0.74) and ODI (0.71). To a great extent for EQ-5D[a] (−0.53).
Item 4: Is the MIC precise? Quantified as width of the 95% CI and expressed as a percentage the estimate	Not so much for analysis 'b' of absolute changes in LBP intensity (38%) (Table S4). Definitely no for all other analyses.(>50%).[b] Precision should ideally be below 10%.	Not so much for analysis 'b' of absolute (35%) and relative changes (27%) in LBP intensity (Table S4). Definitely no for all other analyses (>50%).[b] Precision should ideally be below 10%.
Item 5: Does the threshold or difference between groups on the anchor used to estimate the MIC reflect a small but important difference?	Definitely yes. The seven responses were ''completely recovered'', ''much improved'', ''slightly improved'', ''no change'', ''slightly worsened'', ''much worsened'', and ''worse than ever''. We used values for cutoffs below (ROC-curve analyses), and mean and p25 within, those patients who reported ''much improved'', ''slightly improved''.	Definitely yes. The seven responses were ''completely recovered'', ''much improved'', ''slightly improved'', ''no change'', ''slightly worsened'', ''much worsened'', and ''worse than ever''. We used values for cutoffs below (ROC-curve analyses), and mean and p25 within, those patients who reported ''much improved'', ''slightly improved''.
Additional criteria for transition rating anchors		
Item 1: 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?	Not so much. Three months recall is suboptimal. Follow-up should ideally be sufficiently short for patients to be able to remember their health state at baseline (ideal timing not known, but the credibility instrument suggests <2 months for asserting high credibility).	Definitely no. A recall period of one year is likely too long for patients to be able to remember their health state at baseline (ideal timing not known, but the credibility instrument suggests <2 months for asserting high credibility).
Item 2: Does the transition item have a satisfactory correlation with the PROM score at follow-up?	Definitely yes for all PROMs. Correlations are 0.49 for RMDQ, 0.55 for ODI, 0.65 for LBP intensity NRS and −0.56 for EQ-5D[a].	Definitely yes for all PROMs. Correlations are 0.65 for RMDQ, 0.67 for ODI, 0.76 for LBP intensity NRS and −0.65 for EQ-5D[a].
Item 3: Does the transition item correlate [negatively for RMDQ, ODI and LBP intensity, or positively for EQ-5D] with the PROM score at baseline? Rationale: ''If the score at baseline correlates with the transition rating, we are more confident that patients are taking their baseline status into account when scoring the transition rating'' [3]	Definitely yes for RMDQ (−0.04). To a great extent for ODI (0.06), LBP intensity NRS (0.02) and EQ-5D[a] (−0.05).	Definitely yes for RMDQ (−0.02). To a great extent for LBP intensity NRS (0.09) Not so much for ODI (0.15) and for EQ-5D[a] (−0.16).

(*Continued*)

Table 2. Continued

Signaling questions	Credibility assessment with rationale	
	3-month follow-up	12-month follow-up
Item 4: 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?	Not so much for RMDQ (0.09), ODI (0.05), LBP intensity NRS (0.9) and EQ-5D[a] (−0.01) (numbers refer to difference between correlation GPE—PROM change score and correlation GPE—PROM score at follow-up).	Not so much for RMDQ (0.08), ODI (0.03) and LBP intensity NRS (0.01). Definitely no for EQ-5D[a] (0.12) (numbers refer to difference between correlation GPE—PROM change score and correlation GPE—PROM score at follow-up).
Overall judgment of credibility	We consider the MICs for RMDQ, ODI, LBP intensity NRS and EQ-5D at three-month follow-up to have moderate credibility. The MICs were estimated using a patient-reported anchor that is easily understandable, relevant, with justified cut-offs and were well correlated with change scores in the respective PROMs. However, there is some concern of poor precision, and flawed recall as the follow-up time was at three months. The latter concern was supported by minorly (RMDQ,ODI and LBP intensity NRS) and no (EQ-5D) stronger correlations of the transition item with the PROM change scores compared to the correlations of the transition item with the PROM scores at follow-up. However, the concern was reduced by satisfactory correlations between the transition item and both the baseline and follow-up score.	We consider MICs for RMDQ, ODI, LBP intensity NRS and EQ-5D at 12-month follow-up to have moderate credibility. The MICs were estimated using a patient-reported anchor that is easily understandable, relevant, with justified cut-offs and were well correlated with change scores in the respective PROMs. However, there is a concern of poor precision, and flawed recall as the follow-up time was at one year. The latter concern was supported by minorly (RMDQ,ODI and LBP intensity NRS) and no (EQ-5D) stronger correlations of the transition item with the PROM change scores compared to the correlations of the transition item with the PROM scores at follow-up. However, the concern was not supported as correlations between the transition item and the baseline and the follow-up score of PROM scores were satisfactory. Despite longer follow-up than three-month, correlations between the transition item and the change scores were better for 12-month follow-up.

RMDQ, Roland Morris Disability Questionnaire; ODI, Oswestry Disability Index; LBP (low back pain) intensity, measured by numerical rating scale; EQ5D, Health-related quality of life, EuroQoL-5D; PROM, Patient Reported Outcome Measure.

The credibility assessments are scored according to the instrument for judging the trustworthiness of minimal important differences [3], with grade options yes/no/impossible to tell for item 1 and options definitely yes/to a great extent/definitely no/not so much/impossible to tell for all other items. An overall judgment of credibility (low, moderate, high) is reached based on consideration of the severity of the credibility issue for any particular item and the consequence of this issue. We used the authors of the credibility assessments' own examples of overall judgment (prescriptive approach not available).

[a] For EQ-5D higher values represent different states (worse) than higher values on the anchor global perceived effect (better), which will affect the direction (positive/negative) of the correlations (we used Guide B in the credibility evaluation instrument [3]).

[b] Random-effects logistic regression analyses (combining both follow-up times) for analyses 'b' had better precision: ''To a great extent'' for LBP intensity NRS and ''not so much'' for RMDQ, ODI and EQ-5D.

The sensitivity analyses generally supported the findings in the primary analyses. For analyses 'b' there was a trend of lower MIC estimates with the predictive modeling (that adjusted for the low proportion of improved patients) compared to the ROC-curve analyses (Figure 2). MIC estimates for patients with less symptoms at baseline were often lower than for patients with worse symptoms at baseline when based on analyses of absolute changes, but most often not when based on analyses of relative changes. The predictive modeling that combined information from both timepoints (random-effects logistic regression models, not predefined analyses) gave slightly higher and more precise MIC estimates (Figure S2).

The secondary analyses showed that median improvements in patients who reported to be slightly improved were similar to the MICs obtained from ROC curve analyses 'a' (Figure 1, Table S5). Median improvement in patients who reported much improved were larger than the corresponding MICs obtained from ROC curve analyses 'b'.

We found that the credibility of the MIC estimates were moderate (Table 2 and Table S1).

3.2. Responsiveness

For the ODI and the EQ-5D all hypotheses (6 out of 6) based on a construct approach were confirmed. For RMDQ and LPB intensity NRS five out of six hypotheses were confirmed (Table 3). All PROMs had AUC values > 0.70 for absolute and relative change (criterion approach, Table S3). Responsiveness parameters obtained at the

Table 3. Predefined hypotheses to test responsiveness of PROMs

Hypothesis	Expected value	Calculated value (CI)	Expectation met (hypothesis confimed)
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 [27,28]	rho ≥0.7	0.73 (0.65 to 0.79)	Yes
The correlation between absolute change in RMDQ score and the absolute change in LBP intensity NRS score at 12 months is moderate and positive [27,28]	rho ≥0.3 and < 0.7	0.74 (0.66 to 0.80)	No
The correlation between absolute change in RMDQ score and the absolute change in EQ5D score at 12 months is moderate and positive [29]	rho ≥0.3 and < 0.7	0.61 (0.51 to 0.70)	Yes
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 [27]	SRM <0.2	0.07 (−0.23 to 0.36)	Yes
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 [27]	SRM >0.2	0.82 (0.52 to 1.12)	Yes
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 [27]	SRM >0.5	2.28 (1.72 to 2.83)	Yes
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 [27,28]	rho ≥0.7	0.73 (0.65 to 0.79)	Yes
The correlation between absolute change in ODI score and the absolute change in LBP intensity NRS score at 12 months is moderate and positive [27,28]	rho ≥0.3 and < 0.7	0.67 (0.57 to 0.74)	Yes
The correlation between absolute change in ODI score and the absolute change in EQ5D score at 12 months is moderate and positive [29]	rho ≥0.3 and < 0.7	0.62 (0.51 to 0.70)	Yes
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 [6,27]	SRM <0.2	−0.01 (−0.25 to 0.22)	Yes
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 [6,27]	SRM >0.2	0.95 (0.60 to 1.30)	Yes
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 [6,27]	SRM >0.5	2.11 (1.58 to 2.63)	Yes
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 [27,28]	rho ≥0.3 and < 0.7	0.67 (0.57 to 0.74)	Yes
The correlation between absolute change in LBP intensity NRS score and the absolute change in RMDQ score at 12 months is moderate and positive [27,28]	rho ≥0.3 and < 0.7	0.74 (0.66 to 0.80)	No
The correlation between absolute change in LBP intensity NRS score and the absolute change in EQ5D score at 12 months is moderate and positive [29]	rho ≥0.3 and < 0.7	0.56 (0.45 to 0.66)	Yes
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 [27]	SRM <0.2	0.06 (−0.16 to 0.27)	Yes

(*Continued*)

Table 3. Continued

Hypothesis	Expected value	Calculated value (CI)	Expectation met (hypothesis confimed)
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 [6,27]	SRM >0.2	1.49 (1.05 to 1.93)	Yes
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 [6,27]	SRM >0.5	2.29 (1.74 to 2.85)	Yes
EQ5D			
The correlation between absolute change in EQ5D score and the absolute change in ODI score at 12 months is moderate and positive [27,28]	rho ≥0.3 and < 0.7	0.62 (0.51 to 0.70)	Yes
The correlation between absolute change in EQ5D score and the absolute change in LBP intensity NRS score at 12 months is moderate and positive [27–29]	rho ≥0.3 and < 0.7	0.56 (0.45 to 0.66)	Yes
The correlation between absolute change in EQ5D score and the absolute change in RMDQ score at 12 months is moderate and positive [29]	rho ≥0.3 and < 0.7	0.61 (0.51 to 0.70)	Yes
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 [6]	SRM <0.2	−0.03 (−0.27 to 0.20)	Yes
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 [6]	SRM >0.2	0.56 (0.28 to 0.83)	Yes
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 [6]	SRM >0.5	1.57 (1.11 to 2.03)	Yes

RMDQ, Roland Morris Disability Questionnaire; ODI, Oswestry Disability Index; LBP (low back pain) intensity, measured by numerical rating scale; EQ5D, Health-related quality of life, EuroQoL-5D; SRM, Standardized response mean, the average change divided by the standard deviation of the changes between the paired measurements.

three-month follow-up were comparable to those obtained at 12-month, and parameters obtained for absolute change scores were comparable to those obtained for relative change scores.

4. Discussion

This study found that for PROMs recommended for patients with LBP, the MIC estimates varied depending upon including "slightly improved" in the definition of MIC or not, on using absolute or relative change scores, and on whether adjusting for baseline value of the PROMs or not. The MIC estimates based on relative change scores were more consistent (less dependent on baseline values) compared to absolute change estimates. Our results suggest that MIC is a 15-40% reduction in RMDQ, ODI and LBP intensity NRS scores for chronic LBP; 15-25% when including "slightly improved" in the definition of MIC, and 25-40% without "slightly improved" in the MIC definition. We found moderate credibility of these MIC estimates. The responsiveness for the RMDQ, ODI, LBP intensity NRS and to a lesser extent EQ-5D, was sufficient.

4.1. Strengths and weaknesses

There is a variety of methods and no consensus in the literature on what is the best method for estimating MIC values [30]. Strengths of our study include using updated methods recommended by the COSMIN panel [1], performing predefined, as well as sensitivity analyses, testing assumptions, and assessing credibility.

Limitations of our study include possible weaknesses with the GPE instrument as anchor. First, there could be flawed recall when reporting this instrument at 3- and 12-month follow-up, as was picked up by the credibility assessments (the observed correlations in the additional criteria were not optimal) [31,32]. Secondly, the GPE instrument is domain unspecific, and we cannot know whether patients had one or more specific domains in mind when answering the instrument. Patients might have real improvement in one domain and deterioration in others.

Another limitation is missing data, which could introduce a potential source of bias. Missing not at random cannot be excluded in MIC estimates, and there is no statistical test that properly overcomes this problem [33]. We did however have little missing (<10%) for each individual

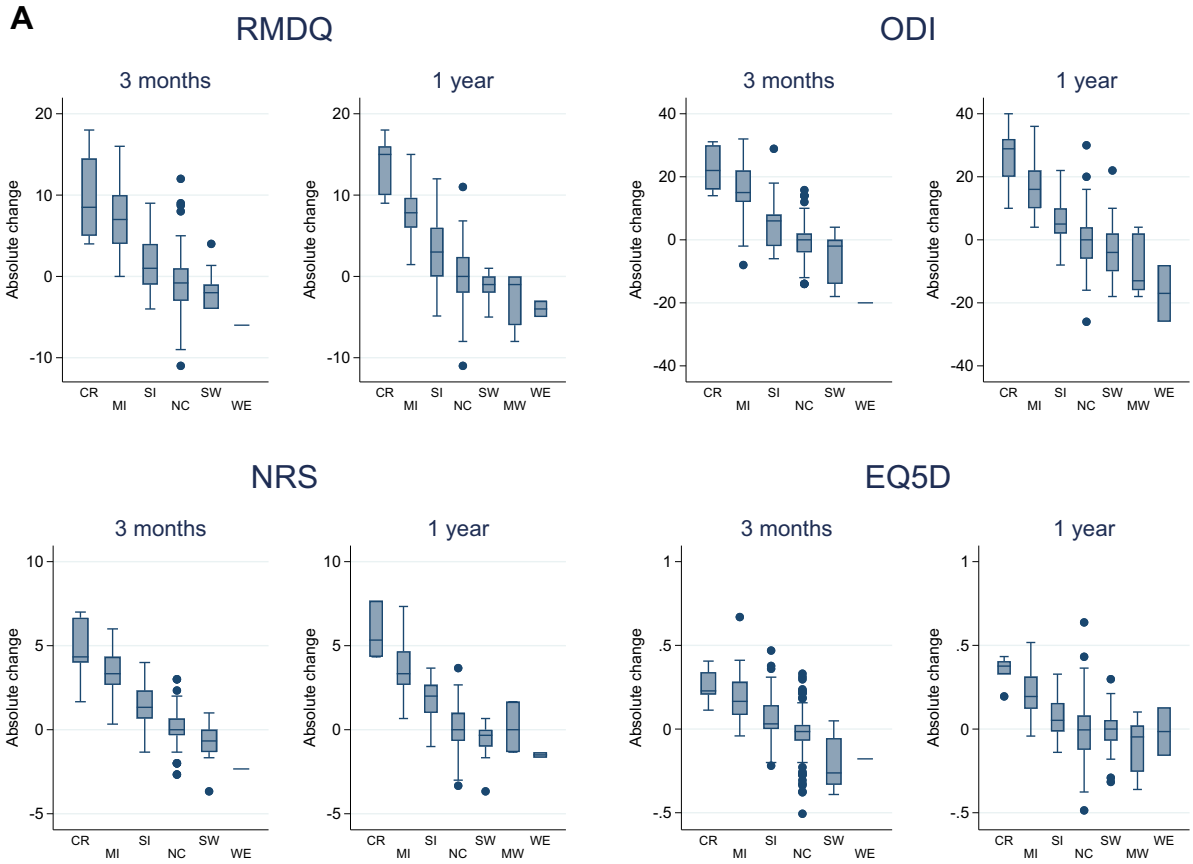

Fig. 1. (A)-Boxplot with descriptive data on absolute change. (B)-Boxplot with descriptive data on relative change. The box plots show median, 25th and 75th percentiles (box) with lower and upper adjacent values (lines outside box) and outliers of absolute (A) and relative (B) change scores for each PROM, within each category of global perceived effect. RMDQ, Roland Morris Disability Questionnaire (0–24); ODI, Oswestry Disability Index (0–100); LBP (low back pain) intensity, measured by numerical rating scale (0–10); EQ5D, Health-related quality of life, EuroQoL-5D (−0.59 to 1); CR, Completely recovered; MI, Much improved; SI, Slightly improved; NC, No change; SW, Slightly worse; MW, Much worse; WE, Worse than ever.

PROM, which makes it unlikely that missing led to high risk of bias. The predictive modeling using random-effects logistic regression analyses (combining data from three- and 12-month follow-up) handles missing better than other analyses, still assuming missing (completely) at random. These analyses suggested slightly larger MIC values compared to the average of MIC estimates from both follow-ups (Figure S2). This slight increase is as expected, as random-effects logistic regression fits subject-specific probabilities that are known to be larger than population-averaged probabilities (estimated in ordinary logistic regression models) [34].

Further, for the analyses 'b' (GPE dichotomized into completely recovered/much improved vs slightly improved/no change) we found a trend of lower estimates from the sensitivity analyses adjusting for the proportion of improved patients compared to the ROC-curve analyses. This is probably explained by the low proportion of improved patients in these analyses, and the predictive modeling might be more accurate than ROC-curve analyses as only the former adjust for this.

4.2. Comparing results with previous studies

Previous MIC estimates are based on studies that used a variety of methods and from a diversity of settings and patient samples [4]. The context should be considered when choosing specific values. The intention of the present study was to provide estimates on a specific patient group (chronic LBP and Modic changes), and comparison to previous results might therefore be less suitable. Some argue against the view that Modic changes represent a specific subgroup, and our results might also be generalizable to a chronic LBP population without Modic changes.

A main determinant of our results was the choice of cut-off on the GPE (analyses 'a' vs 'b'). The MIC estimates for analyses 'a' were on the lower end of the spectrum of estimates presented in previous studies using similar cut-off on the GPE [27,35,36]. We had predefined to consider results from analyses 'a' (15-25% reduction in RMDQ, ODI and LBP intensity scores) relevant for low-risk back pain interventions with few adverse events, and results from the more

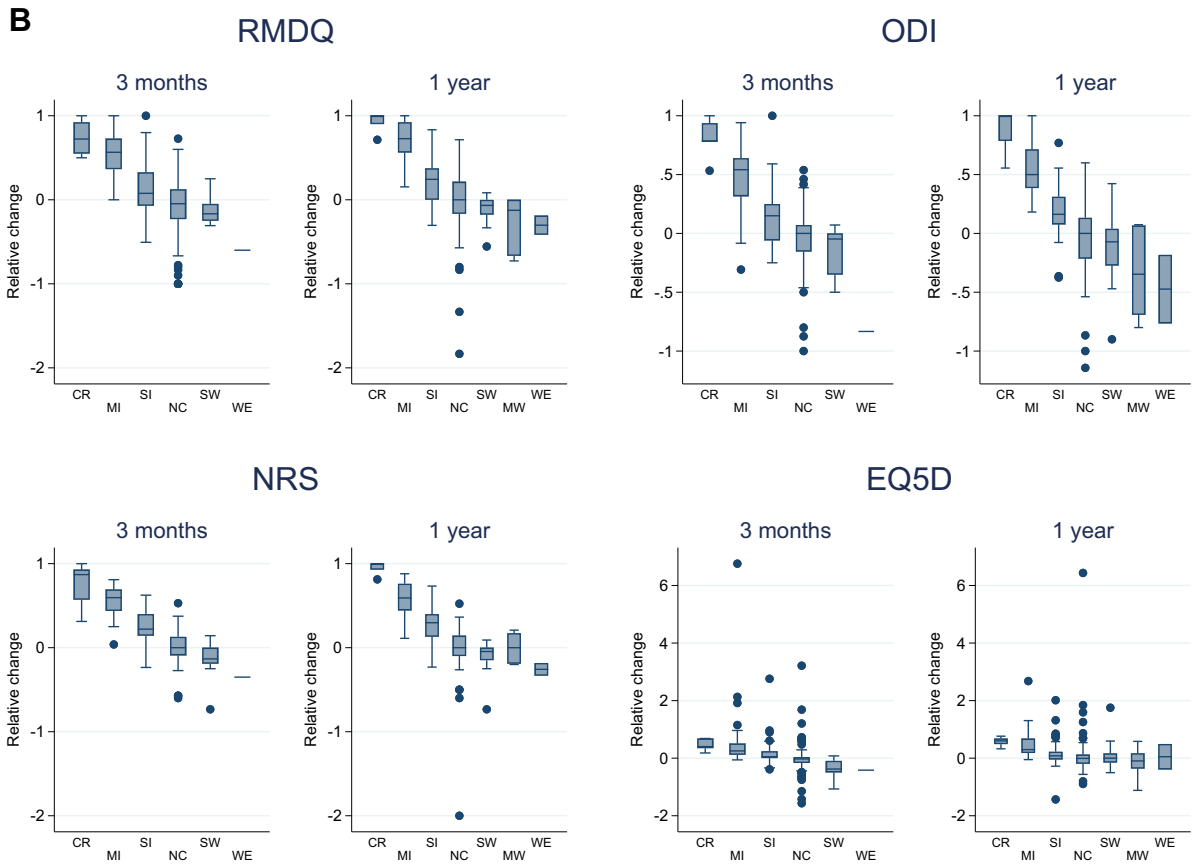

Fig. 1. Continued.

conservative analyses 'b' (25-40% reduction in RMDQ, ODI and LBP intensity scores) relevant for back pain interventions with frequent adverse events (e.g., surgery or antibiotic treatment). Smallest worthwhile effect (between group differences) does indeed depend on risk of adverse events and inconveniences [37], in line with our way of thinking.

For absolute change scores, the MIC-estimates were lower for patients with better compared to patients with worse baseline score. For relative changes in disability and pain intensity, there was less variation between patients with low and high baseline scores. These findings are in line with previous results [38−41] and suggest that MIC estimates of relative changes should be preferred. For EQ-5D it is slightly more complicated to use relative changes as they theoretically can take infinite values (because the denominator can be zero).

4.3. Interpretation of findings

Several points are important when interpreting MIC thresholds (based on ROC curve or predictive modeling analyses) on an individual level. First, one should consider the precision of the cut-point scores/estimates [42]. We want to remind readers that it is wrong to interpret the confidence intervals as the interval that contain 95% of individual MIC thresholds (reference ranges). Sensitivity and specificity express better the extent to which the MIC thresholds applies to individual patients [43]. For analyses 'b', both sensitivity and specificity (not predefined analyses) were mostly better (despite wider confidence intervals of cut-off scores) compared to analyses 'a'. This suggests that estimates from analyses 'b' are more accurate when applied on the individual level. Second, the US FDA guidance suggests that change scores on the individual level should only be viewed as a clinically important change if larger than the smallest detectable change (SDC) [44]. This is defined so that 95% of subjects (with assumed no change in symptoms between two measurements) is expected to report a change in scores less than the SDC, which has reported to be 4 to 9 for RMDQ, 11 to 17 for ODI, 3 to 4 for NRS and 0.4 for EQ5D [9,45,46]. In this respect we can also more readily trust the estimates from analyses 'b', as estimates for analyses 'a' were mostly below values for SDC. However, SDC values for PROMs used in back pain might themselves be uncertain [9].

The credibility assessments of both three- and 12-month follow-ups were similar (both moderate credibility), and we

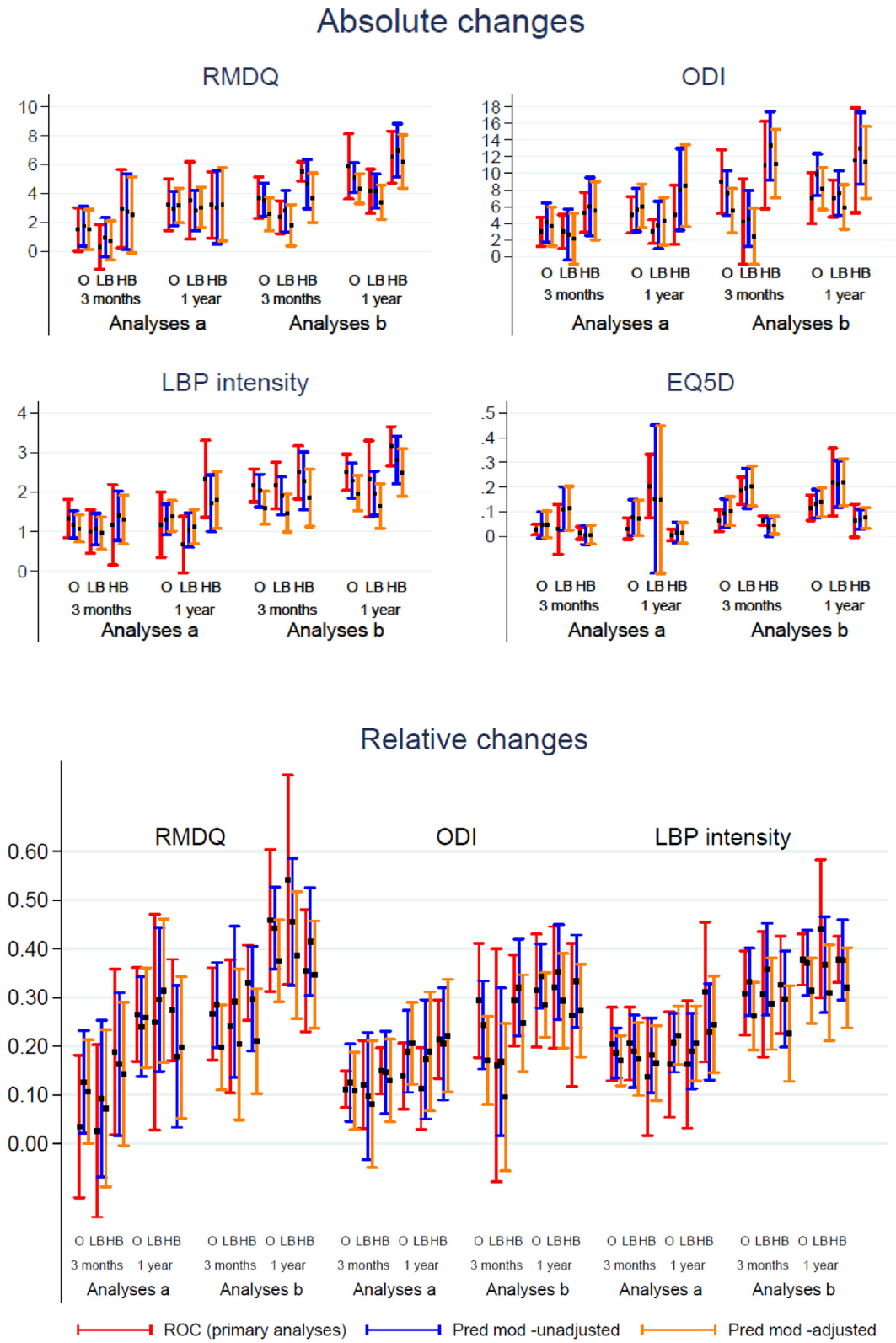

Fig. 2. Minimal important change thresholds analyzed by ROC and Logistic regression for different cutoffs of GPE (analyses 'a' + 'b') Color should be used in print. RMDQ, Roland Morris Disability Questionnaire (0–24); ODI, Oswestry Disability Index (0–100); LBP (low back pain) intensity, measured by numerical rating scale (0–10); EQ5D, Health-related quality of life, EuroQoL-5D (-0.59–1); Analyses a, global perceived effect dichotomized into completely recovered/much improved/slightly improved v no change; Analyses b, global perceived effect dichotomized into completely recovered/much improved vs slightly improved/no change; O, Overall; LB, Low baseline (≤median); HB, High baseline (>median); ROC, Receiver operating curve analyses (Primary analyses, red); Pred mod, Predictive modeling (Sensitivity analyses, blue -unadjusted, orange -

found it reasonable to combine their respective MIC estimates using predictive modeling with random-effects logistic regression models (not predefined analyses). These analyses gave more precise estimates that support our conclusion.

Our estimated within-group MIC thresholds could be used to prespecify a definition of a responder in responder analyses or, with caution due to points mentioned above, to assess individual changes in the clinical setting.

5. Conclusion

This study showed that the MIC varied from 15% to 40% improvement in RMDQ, ODI and LBP intensity scores, values depending on definition of MIC (including those who reported ''slightly improved'' or not). These estimates, tested in a clinical trial of patients with chronic LBP in an outpatient hospital setting, are on the lower spectrum of previous estimates. These core outcomes of back pain showed a sufficient responsiveness.

Acknowledgment

The AIM-study group: University Hospital North Norway, Tromsø (four patients): Terese Fors, Guro Kjos, Ida Beate Østhus (Department of Rehabilitation). Trondheim University Hospital, Trondheim (21 patients): Gunn Hege Marchand, Britt Elin Lurud, Fredrik Granvigen (Department of Physical Medicine and Rehabilitation), Hege Andersen (National Advisory Unit of Spinal Surgery), Øystein Petter Nygaard, Vidar Rao (Department of Neurosurgery). Haukeland University Hospital, Bergen (37 patients): Thomas Istvan Kadar (Department of Physical Medicine and Rehabilitation), Siv Krüger Claussen, Erling Andersen (Department of Clinical Engineering), Per Kristoffersen and Nils Vetti (Department of Radiology), Jörg Aβmus (Center for Clinical Research). Vestre Viken Hospital, Drammen (38 patients): Sigrun Randen (Department of Physical Medicine and Rehabilitation), Hilde Presberg (Department of Neurology). Oslo University Hospital, Oslo (50 patients): Linda Margareth Pedersen, Bendik Slagsvold Winsvold (FORMI), Karianne Wiger Gammelsrud (Department of microbiology), Maria Dehli Vigeland, Benedicte Alexandra Lie, Siri Tennebø Flåm, Magnus Dehli Vigeland (Department of Medical Genetics). Østfold Hospital Trust (30 patients): Marianne Thorsø, Knut Morten Huneide, (Department of Physical Medicine and Rehabilitation), Veronica Sørensen (Department of Rheumatology). Thor Einar Holmgard (patient representative).

We thank Helse Sør-Øst (grant no: 2015090) and Helse Vest (grant no: 911938 and 911891) for funding the AIM-study, KLINBEFORSK (grant no: 2017201) and all patients participating in the study.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2022.07.012.

References

[1] Mokkink LBTC, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, et al. COSMIN checklist manual COSMIN panel. COSMIN; 2012. Available at: http://fac.ksu.edu.sa/sites/default/files/cosmin_checklist_manual_v9.pdf. Accessed August 26, 2022.

[2] Terwee C, Dekker F, Wiersinga W, Prummel M, Bossuyt P. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003;12:349−62.

[3] 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:m1714.

[4] 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.

[5] Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LO, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 2018;159(3):481−95.

[6] 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 Musculoskelet Disord 2019;20(1):31−46.

[7] 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.

[8] Pires D, Cruz E, Canhão H, Nunes CJPT, Practice. Minimum important change values for pain and disability: which is the best to identify a meaningful response in patients with chronic nonspecific low back pain? 2020:1−9.

[9] Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for

adjusted for proportion of improved patients). The table shows that Predictive modeling (sensitivity analyses using logistic regression, in blue and orange) give similar MIC thresholds as the ROC-curve analyses (primary analyses, in red) 'a', but for analyses 'b' there were a trend of lower MIC thresholds when adjusting for the proportion of improved patients in the predictive modeling analyses compared to the unadjusted and the ROC-curve analyses. Further, MIC estimates for patients with better (less symptoms) baseline score were often lower than for patients with worse baseline score based on analyses of absolute changes, but most often not of relative changes. For EQ-5D, low baseline score means worse quality-of-life, and therefore a potential for more improvement than high baseline score. MIC thresholds for relative change of EQ5D are not provided as the scale crosses value zero, and infinite values occur when baseline values are zero. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. Phys Ther 2016;96:1620−37.
[10] Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P, Terwee CB. Measurement properties of visual analogue scale, numeric rating scale, and pain severity subscale of the brief pain inventory in patients with low back pain: a systematic review. J Pain 2019;20(3):245−63.
[11] Bråten LCH, Rolfsen MP, Espeland A, Wigemyr M, Aßmus J, Froholdt A, et al. Efficacy of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM study): double blind, randomised, placebo controlled, multicentre trial. BMJ 2019;367:l5654.
[12] Grotle M, Brox J, Vollestad N. Cross-cultural adaptation of the Norwegian versions of the roland-morris disability questionnaire and the oswestry disability index. J Rehabil Med 2003;35(5):241−7.
[13] Roland M, Fairbank J. The roland-morris disability questionnaire and the oswestry disability questionnaire. Spine 2000;25(24):3115−24.
[14] Fairbank JC, Pynsent PB. The oswestry disability index. Spine 2000; 25(22):2940−52.
[15] 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.
[16] Battie MC, Bigos SJ, Fisher LD, Spengler DM, Hansson TH, Nachemson AL, et al. The role of spinal flexibility in back pain complaints within industry. A prospective study. Spine 1990;15(8):768−73.
[17] Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res 2021;30:2197−218.
[18] Group CSCM. In: Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, editors. Methods to explain the clinical significance of health status measures. Mayo Clinic Proceedings. Rochester, MN: Elsevier; 2002.
[19] Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status. PLoS One 2014;9: e114468.
[20] 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. J Clin Epidemiol 2017;83:90−100.
[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. J Clin Epidemiol 2015; 68:1388−96.
[22] Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002;55:900−8.
[23] Responsiveness and Minimal Important Change for core outcomes in low back pain − Statistical Analysis Plan. AIM study group; 2021. Available at: https://clinicaltrials.gov/ct2/show/NCT02323412?cond=modic+change&rank=1. Accessed August 26, 2022.
[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. J Clin Epidemiol 2007;60:34−42.
[25] de Vet HTC, Mokkink L, Knol D. Measurement in Medicine Cambridge. Cambridge: Cambridge University Press; 2011.
[26] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29−36.
[27] 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):E492−501.
[28] 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 Qual Life Outcomes 2020;18(1):1−12.
[29] Soer R, Reneman MF, Speijer BL, Coppes MH, Vroomen PC. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. Spine J 2012;12(11):1035−9.
[30] Devji T, Carrasco-Labra A, Guyatt G. Mind the methods of determining minimal important differences: three critical issues to consider. Evidence-Based Ment Health 2021;24(2):77−81.
[31] 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. J Clin Epidemiol 2010;63:760−766.e1.
[32] 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. J Clin Epidemiol 2014;67:508−15.
[33] Woaye-Hune P, Hardouin J-B, Lehur P-A, Meurette G, Vanier A. Practical issues encountered while determining minimal clinically important difference in patient-reported outcomes. Health Qual Life Outcomes 2020;18:1−13.
[34] Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using Stata. 3rd ed. College Station, TX: Stata Press; 2012.
[35] Van Der Roer N, Ostelo RW, Bekkering GE, Van Tulder MW, De Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine 2006;31(5):578−82.
[36] Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol 2006;59:45−52.
[37] Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Grotle M, et al. The smallest worthwhile effect of nonsteroidal antiinflammatory drugs and physiotherapy for chronic low back pain: a benefit-harm trade-off study. J Clin Epidemiol 2013;66:1397−404.
[38] de Vet HC, Foumani M, Scholten MA, Jacobs WC, Stiggelbout AM, Knol DL, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. J Clin Epidemiol 2015;68:518−24.
[39] Wang Y-C, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. Phys Ther 2011;91:675−88.
[40] Chiarotto A, Vanti C, Cedraschi C, Ferrari S, Ostelo RW, Pillastrini P. Responsiveness and minimal important change of the pain self-efficacy questionnaire and short forms in patients with chronic low back pain. J Pain 2016;17(6):707−18.
[41] Olsen MF, Bjerre E, Hansen MD, Hilden J, Landler NE, Tendal B, et al. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. BMC Med 2017;15(1):1−18.
[42] Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. Otolaryngol Head Neck Surg 2019;161(4):551−60.
[43] de Vet HCW, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RWJG, et al. Three ways to quantify uncertainty in individually applied ''minimally important change'' values. J Clin Epidemiol 2010;63:37−45.
[44] McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev pharmacoecon Outcomes Res 2011;11(2):163−9.
[45] 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. J Med Internet Res 2018;20(10):e272.
[46] Johnsen LG, Hellum C, Nygaard ØP, Storheim K, Brox JI, Rossvoll I, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord 2013;14(1):1−9.