Roland-Morris Disability Questionnaire, Oswestry Disability Index, and Quebec Back Pain Disability Scale

Jenks, Alan; Hoekstra, Trynke; van Tulder, Maurits; Ostelo, Raymond W.; Rubinstein, Sidney M.; Chiarotto, Alessandro

published in
Journal of Orthopaedic and Sports Physical Therapy
2022

DOI (link to publisher)
10.2519/jospt.2022.10802

document version
Publisher's PDF, also known as Version of record

ALAN JENKS, DC[1,2] • **TRYNKE HOEKSTRA,** PhD[1,3] • **MAURITS VAN TULDER,** PhD[2,6,7] • **RAYMOND W. OSTELO,** PT, PhD[1,2,4]
SIDNEY M. RUBINSTEIN, DC, PhD[1,2] • **ALESSANDRO CHIAROTTO,** PT, PhD[1,2,5]

Roland-Morris Disability Questionnaire, Oswestry Disability Index, and Quebec Back Pain Disability Scale: Which Has Superior Measurement Properties in Older Adults With Low Back Pain?

OBJECTIVE: To examine the validity, reliability, and responsiveness of 3 commonly used questionnaires for assessing physical function (ie, Oswestry Disability Index [ODI], Quebec Back Pain Disability Scale [QBPDS], and Roland-Morris Disability Questionnaire [RMDQ]) in older patients undergoing chiropractic care for low back pain (LBP).

DESIGN: Head-to-head clinimetric comparison.

METHODS: Patients completed the ODI, QBPDS, and RMDQ at baseline and after 2 weeks of treatment. Reliability was evaluated for internal consistency (Cronbach α), test-retest reliability (interclass correlation coefficient [ICC]), and measurement error (standard error of measurement and smallest detectable change [SDC]). Structural validity was evaluated through unidimensional confirmatory factor analysis, and construct validity was investigated by a priori hypotheses with other measures. Responsiveness was evaluated by testing a priori hypotheses using data at baseline and at 2-week follow-up.

RESULTS: Two hundred fourteen patients (53% males and 47% females) with a mean age of 66.2 years (standard deviation = 7.8 years) were included, of which 193 patients completed the 2-week follow-up for our responsiveness analysis. The RMDQ, ODI, and QBPDS showed sufficient internal consistency (Cronbach α of .89, .86, and .94, respectively) and test-retest reliability (ICC[2,1] of 0.85, 0.89, and 0.84, respectively). The SDC for the RMDQ was 6.9, for the ODI was 19.1, and for the QBPDS was 23.6, which are values larger than the minimal important change. None of the measures met all criteria for sufficient structural validity, but the RMDQ and ODI exhibited a partial unidimensional fit. The questionnaires had sufficient construct validity and responsiveness.

CONCLUSION: The ODI, QBPDS, and RMDQ have similar measurement properties in older adults with LBP. *J Orthop Sports Phys Ther 2022;52(7):457-469. Epub: 18 May 2022. doi:10.2519/jospt.2022.10802*

KEY WORDS: *anatomy/spine, clinical measurement (clinimetrics), low back/lumbar spine, manual therapy/spine, outcome measures*

There is a high prevalence of low back pain (LBP) in older adults[11]—a musculoskeletal problem that is not well understood[26,41] or treated. Given multiple factors (eg, psychological and physical comorbidities, maladaptive coping, and age-related physical problems) can modify the LBP experience in older adults.[47,50]

Physical function is a core outcome domain for patients with LBP.[6,24,43,51] International consensus recommends using the Roland-Morris Disability Questionnaire (RMDQ) or the Oswestry Disability Index (ODI) to measure physical function in clinical trials.[5] They are the most frequently used patient-reported outcome measures (PROMs) for physical function in the adult population.[19] Another common measure is the Quebec Back Pain Disability Scale (QBPDS),[4] which has promising measurement properties.[8]

A measurement instrument needs adequate measurement properties (ie, validity, reliability, and responsiveness).[30] Instruments must be evaluated in head-to-head studies, where they are administered to the same target population, in the same setting, at the same time points, and with the same comparator instruments.[7] In recent systematic reviews of head-to-head comparisons, there was no single instrument (RMDQ, ODI, or

[1]Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2]Amsterdam Movement Sciences Research Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [3]Amsterdam Public Health Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [4]Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Location VUmc, Amsterdam, the Netherlands. [5]Department of General Practice, Erasmus University Medical Center, Rotterdam, the Netherlands. [6]Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [7]Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark. This protocol has received ethical approval from the Medical Ethics Committee of the VU University Medical Center (ethics number 2017-618). Written informed consent was obtained from the patient for the publication of their individual details and accompanying images in this manuscript. The consent form is held by the authors' institution in a secured server and is available for review by the editor in chief. This study is funded by the European Centre for Chiropractic Research Excellence (ECCRE), located in Odense, Denmark (grant number 01-2016-NL/MvT), and by the Nederlandse Chiropractoren Associatie (NCA), located in Joure, the Netherlands. Drs Jenks and Rubinstein work as chiropractors in private practice. Drs van Tulder and Rubinstein received grants from the European Chiropractors' Union, the ECCRE, the Belgian Chiropractic Association (BVC), and the NCA for their positions at the Vrije Universiteit Amsterdam. The other authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Alan Jenks, Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, MF Building, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands. E-mail: a.d.jenks@vu.nl

QBPDS) that came out on top in terms of measurement properties.[7,32] All failed in some key measurement aspects[8]: issues with the unidimensionality of the total score (which is routinely used) and with measurement error (measured as the smallest detectable change [SDC]), which is usually larger than 20% of the score range.[7,8,44]

Few studies have focused on the measurement properties of the RMDQ, ODI, and QBPDS in older adults (older than 65 years) with LBP. Hicks and Manal[21] found sufficient test-retest reliability and convergent/construct validity of the ODI and QBPDS in older adults but did not evaluate responsiveness. Davidson and Keating[10] examined reliability and responsiveness in older adults with LBP but did not evaluate validity. Additionally, the sample sizes were too small to draw any firm conclusion.[10] Most importantly, no single study evaluated a head-to-head analysis in the 3 domains of validity, reliability, and responsiveness.

We aimed to compare the validity, reliability, and responsiveness of the RMDQ, ODI, and QBPDS in older adults with LBP in a head-to-head clinimetric comparison.

METHODS

This clinimetric study used data from a prospective observational study[26] with measurements at baseline and after 2 weeks. Chiropractors from The Netherlands Chiropractic Association were asked to participate. The chiropractors who agreed were spread across the Netherlands and recruited patients from their practices between September 2018 and December 2019. Patients who called the practices to book an appointment and who met the inclusion criteria were invited to participate.

The data are part of a larger international cohort of the BAck Complaints in the Elders – Chiropractic (BACE-C)[26] study. Patients were eligible if they were aged 55 years or older, had LBP (with or without leg symptoms), and had not seen a chiropractor in the previous 6 months. LBP was defined as pain from the thoracolumbar 12th rib junction to the first sacral vertebrae, including pelvic pain and pain referred to the leg. Exclusion criteria were as follows: inadequate command of the Dutch language and no Internet access via a computer, tablet, or smartphone. We excluded people with a cognitive disorder and those with suspected tumor, fracture, infection, or any other potential red flag or condition considered a contraindication for spinal manipulation therapy. The ethics committee of the VU University Medical Center approved the study protocol (2017-618). All patients gave online informed consent to participate in the study.

Data Collection

A link to the questionnaire was e-mailed to patients and completed as a web-based questionnaire at baseline and after 2 weeks. A 2-week time interval was chosen for follow-up based on previous research,[7,18] as we expected little change within a 2-week period, and it would capture more patients as stable. During baseline and 2-week data collection, participants received chiropractic care based on the chiropractor's pragmatic treatment plan. Patients received treatment at least once a week, and frequency varied between once and 3 times a week. We did not record the duration of treatment or the specifics of what each treatment consisted of as the chiropractor was free to treat based on clinical need.

Baseline questionnaires captured the following: (1) sociodemographic characteristics (eg, age, sex, marital status), (2) physical activity (measured with the International Physical Activity Questionnaire [IPAQ]),[2] (3) LBP information including an 11-point numeric rating scale to measure pain intensity as well as items on duration of pain and onset and previous episodes of LBP, (4) the Dutch version of the 3 physical functioning PROMs (ie, 24-item RMDQ,[3,37] ODI Version 2.1a,[16,48] and QBPDS[28,42]), (5) health-related quality of life measured with the EQ-5D-5L (EuroQol 5 Dimension 5 Level),[49] (6) comorbidities using the Self-Administered Comorbidity Questionnaire,[45] and (7) the STarT Back Screening Tool.[40]

At the 2-week follow-up, a 7-point "global change" scale was included with the questionnaires. Participants rated the extent to which their back problem had changed from the start of treatment. The rating scale had 7 response options: 1 = a lot better, 2 = much better, 3 = better, 4 = a little better, 5 = about the same, 6 = a little worse, and 7 = much worse. All 3 PROMs, the IPAQ, and the EQ-5D-5L questionnaires were completed at the 2-week follow up.

Physical Functioning PROMs

The 24-item RMDQ[21,37] was developed in 1983 using and modifying items from the Sickness Impact Profile. It consists of 24 items that represent activities routinely done or avoided that are likely affected by LBP. A 0-to-24 sum score is calculated by counting the number of endorsed items,[37] with higher scores indicating worse function. The RMDQ score used a modified proportional recalculation method[19] to deal with missing data. The online version of the RMDQ was a forced submission; there were no missing data.

The ODI Version 1.0[16,48] was first published in 1980 as an indicator of disability (defined as the limitation of a patient's performance compared with that of a fit person). The ODI consists of 10 items representing different health constructs. The sum of the section scores is divided by the total possible score (50 if all sections are completed), and the total is multiplied by 100 to yield a percentage score. We used the ODI Version 2.1a.[36] Missing data from the ODI were considered in the total score calculation by removing the question from the total score and recalculating the total score based on completed items.

The QBPDS[28,42] was published in 1995. It provided a scale based on a conceptual model via interviews with patients using item response theory.[27] The authors aimed

to develop a scale that was sufficiently informative over a wide range of disability levels and responsive.[28] The total score is calculated by adding the 20 individual item scores; each item score ranges from 0 to 5, and the total score ranges from 0 to 100. The online version of the QBPDS was a forced submission; there were no missing data.

In all 3 PROMs, a lower score after baseline indicated better physical function.

Measurement Properties and Statistical Methods

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy[30] was used to define the measurement properties under investigation. Statistical analyses were performed using SPSS Version 26. Mplus Version 8.4 was used for the confirmatory factor analysis.

Structural and Construct Validity

To evaluate structural validity, a confirmatory factor analysis for a single-factor solution[17,20,48,52] was performed considering that the total scores of these questionnaires are routinely used. To evaluate model fit and unidimensionality, the comparative fit index (CFI), Tucker-Lewis index (TLI), root-mean-square error of approximation (RMSEA), and standardized root-mean-square residual (SRMR) were used. Guidelines[23] suggest that a CFI and TLI of 0.95 or higher, an RMSEA close to or below 0.06, and an SRMR close to 0.08 or higher represent good fitting models. An instrument was deemed to have better structural validity if it fulfilled all these criteria.

Construct validity was evaluated by testing 7 a priori specified hypotheses (**TABLE 1**). These hypotheses were formulated based on previous research[9,29,31] as well as discussion and agreement among 4 researchers (A.J., T.H., S.M.R., and A.C.). For sufficient construct validity, an instrument was required to meet at least 75% of these hypotheses.[34] The rationale for the hypothesis is presented in **TABLE 1**.

Reliability

Reliability was evaluated by testing (1) internal consistency, (2) test-retest reliability, and (3) measurement error.

The Cronbach α was used as a parameter of internal consistency based on the total score. A Cronbach α coefficient greater than or equal to .70 and less than .95 is regarded as satisfactory.[34] The Cronbach α if item deleted was also calculated.

Test-retest reliability was explored in the subgroup of patients identified post hoc as stable based on self-reporting of their condition as "a little better," "about the same," or "a little worse" on the 7-point global change scale. Interclass correlation coefficients for agreement ($ICC_{agreement}$) were calculated for each instrument. A PROM was considered to have better test-retest reliability if displaying an $ICC_{agreement}$ greater than or equal to 0.70 or an $ICC_{agreement}$ of at least 0.10 higher than another PROM.[30] A sensitivity analysis categorizing "a little better" as stable was also explored to confirm the robustness of the results.

We defined measurement error by calculating the SDC, which was SEM $\times$ 1.96 $\times$ $\sqrt{2}$. First, we calculated the standard error of measurement (SEM) for agreement and then the error associated with repeated measures to identify the SDC. A PROM had sufficient measurement error if the SDC was less than the minimal important change (MIC).[34] The MIC values proposed by Ostelo et al[33] were used to assess this measurement property (ie, 5 for the RMDQ, 10 for the ODI, and 20 for the QBPDS) (**TABLE 5**). Percent in scale range (RMDQ, 28.6; ODI, 19.1; QBPDS, 23.6) was calculated by dividing the SDC by the total range score of the instrument and converting it into a percentage.

Responsiveness

We examined the ability of the RMDQ, ODI, and QBPDS to detect change over 2 weeks. Five hypotheses (**TABLE 1**) were formulated regarding expected mean differences between change scores of the instruments, on expected correlations between changes in scores on the instruments. The hypotheses were formulated based upon previous research[9,29,31] and consensus. Standardized mean responses were calculated by dividing mean change scores by the respective standard deviations (SDs) of the change. The area under the curve (AUC) was calculated as the probability of correctly discriminating patients as stable or unchanged. An AUC greater than 0.70 and close to 0.94 was considered sufficient.[34] Acceptable responsiveness requires that more than 75% of the hypotheses be confirmed[34] (**TABLE 1**).

RESULTS

A total of 286 patients were eligible for the study, of which 214 (75%) patients agreed to participate and completed the baseline questionnaire. There were 193 (90%) patients who completed follow-up measures at 2 weeks. The mean age at baseline was 66 years (SD = 7.8), 47% were female, and the median duration of LBP was 214 days (interquartile range, 2 days to 31 years) (**TABLE 2**). The mean score at baseline of the RMDQ was 9.8 (SD = 5.5) (score, 0-24) or 41 on a 100-point scale, the ODI was 23.0 (SD = 16.3) (score, 0-100), and the QBPDS was 31.6 (SD = 17.6) (score, 0-100).

A group of participants who did not wish to participate in the follow-up measurements completed a paper version of the baseline questionnaire. The baseline descriptive variables were similar in both those who participated and those who did not participate in the follow-up measurements (age, 70 [8.2], [57-83]; 47% females, 53% males; body mass index, 28.3 [6.5]), indicating no selection bias. Electronic versions of physical functioning are adequate for clinical and research settings for assessing patients with chronic LBP.[1]

We classified 110 (51%) patients as "Not Stable" and 83 (39%) as "Stable" (**TABLE 3**). Twenty-one (10%) respondents

TABLE 1 A Priori Hypotheses to Assess RMDQ, ODI, and QBPDS Construct Validity and Responsiveness in Older Patients With Chronic Low Back Pain (n = 193)

	Hypotheses for Construct Validity[a]	RMDQ		ODI		QBPDS	
1.	**Demographic**						
	The mean score (standard deviation) of the back-specific PROMs in people aged >75 is higher than the mean score of the back-specific PROMs in people aged <75.	>75 9.9 (5.5)	<75 6.5 (5.1)	>75 22.7 (16.5)	<75 15.8 (14.6)	>75 31.8 (17.9)	<75 21.9 (16.5)
	https://doi.org/10.1111/j.1526-4637.2003.03042.x	**Confirmed +**		**Confirmed +**		**Confirmed +**	
2.	The mean score (standard deviation) of the back-specific PROMs in people with lower education is higher than the mean score of the back-specific PROMs in people with higher education.	↓ Educ 8.7 (0.57)	↑ Educ 8.0 (0.53)	Educ 23.1 (1.7)	Educ 19.2 (1.6)	Educ 30.3 (1.8)	Educ 26.4 (1.7)
	https://doi.org/10.1186/1471-2474-15-255	**Confirmed +**		**Confirmed +**		**Confirmed +**	
3.	**Physical Tests**						
	The mean score (standard deviation) of the back-specific PROMs in people walking less than 200 m is higher than the mean score of the back-specific PROMs in people walking more than 200 m.	<200 m 13.3 (4.7)	>200 m 7.3 (4.4)	<200 m 37.7 (17.5)	>200 m 16.8 (11.9)	<200 m 45.0 (14)	>200 m 24.1 (13.8)
	https://dx.doi.org/10.3390%2Fjcm9041023	**Confirmed +**		**Confirmed +**		**Confirmed +**	
4.	**Pain**						
	The correlation between the back-specific PROMs and the NRS pain scale is at least 0.1 higher than the correlation between the back-specific PROMs and the EQ-5D-5L anxiety domain.	NRS 0.44	EQ-5D-5L 0.26	NRS 0.41	EQ-5D-5L 0.22	NRS 0.41	EQ-5D-5L 0.23
	http://dx.doi.org/10.1016/j.spinee.2012.10.030	**Confirmed +**		**Confirmed +**		**Confirmed +**	
5.	**Quality of Life**						
	The correlation between the back-specific PROMs and the EQ-5D-5L mobility subscale is 0.2 higher than the correlation between the back-specific PROMs and the EQ-5D-5L anxiety subscale.	Mobility 0.57	Anxiety 0.26	Mobility 0.56	Anxiety 0.22	Mobility 0.53	Anxiety 0.23
	http://dx.doi.org/10.1016/j.spinee.2012.10.030	**Confirmed +**		**Confirmed +**		**Confirmed +**	
6.	**Physical Activity**						
	The mean score (standard deviation) of the back-specific PROMs in people with expected decrease in activity over the next 3 months is higher than the mean score of the back-specific PROMs in people with no expected decrease in activity over the next 3 months.	<Activity 10.6 (5.3)	Activity 8.4 (5.2)	<Activity 24.1 (16)	Activity 21.3 (17.1)	<Activity 33.4 (18)	Activity 28.2 (16.4)
	https://dx.doi.org/10.3390%2Fjcm9041023	**Confirmed +**		**Confirmed +**		**Confirmed +**	
7.	**Worrisome**						
	The correlation between the back-specific PROMs and the STarT Back Tool score is 0.2 higher than the correlation between the back-specific PROMs and the STarT Back distress subscale.	STarT 0.63	Distress 0.53	STarT 0.61	Distress 0.53	STarT 0.46	Distress 0.42
	http://dx.doi.org/10.1136/bmjopen-2016-012445	**Not confirmed –**		**Not confirmed –**		**Not confirmed –**	
	Number That Met Hypotheses (%)	83%		83%		83%	

Table continues on next page.

TABLE 1 A Priori Hypotheses to Assess RMDQ, ODI, and QBPDS Construct Validity and Responsiveness in Older Patients With Chronic Low Back Pain (n = 193) (continued)

	Hypotheses for Responsiveness[a]	RMDQ		ODI		QBPDS	
1.	The correlation of the change scores on the back-specific PROMs with the change scores on the NRS pain scale is 0.2 higher than the correlation of the change scores on the back-specific PROMs with the STarT Back Tool distress subscale high-risk group.	NRS	Distress	NRS	Distress	NRS	Distress
		0.43[b]	0.18[b]	0.34[b]	0.3[b]	0.42[b]	0.09[b]
	http://dx.doi.org/10.1136/bmjopen-2016-012445	**Confirmed +**		**Not confirmed –**		**Confirmed +**	
2.	The correlation of the change scores of the back-specific PROMs and the NRS change scores is 0.2 higher than the correlation of the change scores of the back-specific PROMs and IPAQ change scores.	NRS	IPAQ	NRS	IPAQ	NRS	IPAQ
		0.35[b]	–0.08	0.29[b]	0.04	0.39[b]	–0.02
	https://doi.org/10.1590/S1980-6574201700020015	**Confirmed +**		**Confirmed +**		**Confirmed +**	
3.	The standardized mean response (standard deviation) of the back-specific PROMs in patients categorized with acute low back pain (<6 weeks) is larger than the standardized mean response of the back-specific PROMs in patients categorized with chronic low back pain (>12 weeks).	Acute	Chronic	Acute	Chronic	Acute	Chronic
		8.6 (4.7)	7.6 (5.0)	20.4 (13.5)	21.3 (15.5)	28.5 (14.2)	26.2(16.8)
	http://dx.doi.org/10.6061/clinics/2019/e789	**Confirmed +**		**Not confirmed –**		**Confirmed +**	
4.	The standardized mean response (standard deviation) of the back-specific PROMs in patients categorized as "improved" on the GPE is at least larger than the standardized mean response of the back-specific PROMs in patients categorized as "not improved" on the GPE.	Improved	Not improved	Improved	Not improved	Improved	Not improved
		7.5 (4.7)	8.4 (4.8)	17.1 (12.3)	20.4 (15.2)	24.1 (14.5)	28.2 (15.7)
	https://doi.org/10.1093/ageing/afw127	**Confirmed +**		**Confirmed +**		**Confirmed +**	
5.	The area under the curve is higher than 0.70 for the change scores of the PROMs in patients categorized as "improved" on the GPE.	0.75		0.72		0.75	
		Confirmed +		**Confirmed +**		**Confirmed +**	
	Number That Met Hypotheses (%)	100%		80%		100%	

Abbreviations: Educ, education; EQ-5D-5L, EuroQol 5 Dimension 5 Level; GPE, Global Perceived Effect scale; IPAQ, International Physical Activity Questionnaire; NRS, numeric rating scale; ODI, Oswestry Disability Index; PROMs, patient-reported outcome measures; QBPDS, Quebec Back Pain Disability Scale; RMDQ, Roland-Morris Disability Questionnaire.

[a]*The DOIs provided are research supporting the hypothesis.*

[b]*Correlations were statistically significant: P<.000.*

did not answer the 7-point global change follow-up question.

Validity

Confirmatory factor analysis (**TABLE 4**) suggested that the 1-factor solution did not adequately fit the QBPDS (CFI, 0.88; TLI, 0.87; RMSEA, 0.18; SRMR, 0.09). There was a partial fit for the RMDQ (CFI, 0.93; TLI, 0.93; RMSEA, 0.06; SRMR, 0.12) and the ODI (CFI, 0.96; TLI, 0.95; RMSEA, 0.07; SRMR, 0.06). The ODI had 9 missing iterations due to many patients skipping question 8 (asking about LBP during sex). The RMDQ and QBPDS had no missing iterations. A confirmatory factor analysis was performed for 23 of 24 questions of the RMDQ. Question 24 was not discriminative enough as all participants responded "no" to the question "I stay in bed most of the time because of my back."

The results of hypothesis testing indicated that all 3 questionnaires had sufficient construct validity—met at least 75% of the hypotheses (**TABLE 1**).

Reliability

Cronbach α values at baseline for the RMDQ (.89), ODI (.86), and QBPDS (.94) indicated sufficient internal consistency. Item deletion had no change in the results of the Cronbach α (**APPENDIX**).

TABLE 2 Baseline Descriptives

	Baseline	2 Weeks	Excluded Participants (Baseline Only)	Missing at Baseline
Demographic data	n = 214	n = 193	n = 21	
Age, mean (SD) [IQR], years	66.2 (7.8) [55-96]	66.3 (7.8) [55-96]	65.3 (7.5) [55-78]	0 (0%)
Sex				
Female, n (%)	100 (47%)	94 (47%)	8 (50%)	0 (0%)
Male, n (%)	114 (53%)	107 (53%)	8 (50%)	0 (0%)
Body mass index, mean (SD)	26.2 (4.2)	26.2 (4.3)	28.7 (5.6)	0 (0%)
Lifestyle factors				
Physical activity, median (range), min/week	660 (0-1980)	668 (0-1860)	583 (0-960)	
Smoker				
Yes, n (%)	24 (11%)	19 (9%)	2 (13%)	
No, n (%)	171 (80%)	167 (83%)	12 (75%)	19 (9%)
Alcohol consumption				
Never, n (%)	26 (12%)	25 (12%)	1 (6%)	
1-3× per month, n (%)	106 (50%)	99 (49%)	10 (62%)	
4× or more per month, n (%)	63 (29%)	62 (31%)	3 (19%)	19 (9%)
Back pain with sleeping				
Never, n (%)	56 (27%)	55 (26%)	5 (32%)	
<1-2× per week, n (%)	101 (47%)	99 (73%)	2 (12%)	
3× per week or more, n (%)	52 (24%)	57 (27%)	9 (56%)	5 (2%)
Sociodemographics				
Ethnicity, n (%), Dutch	204 (95%)	196 (97%)	16 (100%)	0 (0%)
Marital status				
Single, n (%)	29 (14%)	29 (14%)		
Married, n (%)	178 (83%)	165 (82%)	16 (100%)	
Living apart together, n (%)	7 (3%)	7 (4%)		0 (0%)
Level of education				
Low, n (%)	33 (15%)	30 (15%)	4 (25%)	
Middle, n (%)	100 (46%)	95 (47%)	10 (62%)	
High, n (%)	81 (38%)	76 (38%)	2 (13%)	0 (0%)
Employment status				
At work, n (%)	77 (36%)	72 (36%)	7 (49%)	
Not at work, n (%)	137 (65%)	125 (62%)	7 (49%)	0 (0%)
Nature and severity of LBP				
Combined pain scores at this moment, mean (SD)	5.5 (2.2)	5.5 (2.2)	10.9 (12.1)	0 (0%)
Combined pain scores this past week, mean (SD)	5.9 (2.2)	6.0 (2.1)	6.0 (2.5)	0 (0%)
Combined pain scores at this moment in your leg, mean (SD)	4.3 (2.5)	4.3 (2.5)	5.8 (2.5)	0 (0%)
Combined pain scores this past week in your leg, mean (SD)	5.4 (2.4)	4.5 (9.9)	5.9 (2.4)	0 (0%)
Previous episode				
Yes, n (%)	173 (81%)	143 (81%)	12 (75%)	
No, n (%)	41 (19%)	38 (19%)	4 (25%)	0 (0%)
Started with a bad movement, n (%)	45 (21%)	40 (20%)	5 (31%)	
Started with heavy lifting, n (%)	12 (6%)	10 (5%)		
Accident/trauma, n (%)	9 (4%)	8 (4%)		
Start slowly over days, n (%)	36 (17%)	35 (17%)	4 (25%)	

Table continues on next page.

TABLE 2 Baseline Descriptives (continued)

	Baseline	2 Weeks	Excluded Participants (Baseline Only)	Missing at Baseline
Other*, n (%)	112 (52%)	108 (54%)	7 (44%)	1 (1%)
Duration of low back pain, n = mean # days (IQR)	51 (4-279)			
Frequency of low back pain				
<1× per week, n (%)	19 (9%)	16 (8%)	3 (19%)	
1× per week, n (%)	4 (2%)	2 (1%)	1 (6%)	
Every day, n (%)	171 (80%)	165 (82%)	7 (44%)	
Every minute of the day, n (%)	20 (9%)	18 (9%)	5 (31%)	0 (0%)
Pain referral to leg				
Yes, n (%)	125 (58%)	115 (57%)	13 (81%)	
No, n (%)	87 (41%)	84 (42%)	3 (19%)	0 (0%)
Numbness or tingling in leg or foot				
None-mild, n (%)	169 (92%)	161 (80%)	8 (50%)	
Moderate-very severe, n (%)	44 (8%)	67 (33%)	8 (50%)	0 (0%)
Weak or heavy feeling in leg or foot				
None-mild, n (%)	152 (71%)			
Moderate-very severe, n (%)	62 (29%)			0 (0%)
Average max walking in the last week				
More than 3 km, n (%)	97 (46%)	91 (45%)	6 (38%)	
200 m-3 km, n (%)	89 (42%)	84 (42%)	6 (38%)	
15 m-200 m, n (%)	22 (10%)	20 (10%)	4 (24%)	
Less than 15 m, n (%)	6 (3%)	6 (3%)		0 (0%)
Present episode of pain				
Comes and goes, n (%)	124 (58%)	115 (57%)	6 (38%)	
Constant, n (%)	90 (42%)	86 (43%)	10 (62%)	0 (0%)
Expectations of treatment				
Recovery/improvement, n (%)	200 (94%)	193 (96%)	14 (88%)	
Stay about the same, n (%)	11 (5%)	7 (4%)	2 (12%)	2(1%)
Expectations of recovery after 3 months				
Pain free, n (%)	59 (28%)	53 (26%)	8 (50%)	
Large improvement, n (%)	137 (64%)	134 (66%)	4 (25%)	
About the same, n (%)	18 (8%)	14 (7%)	4 (25%)	13 (6%)
Expectations of work/activity in next 3 months				
Fully recovered, n (%)	92 (43%)	90 (45%)	8 (50%)	
Partially recovered, n (%)	23 (11%)	21 (10%)	4 (25%)	
About the same, n (%)	21 (10%)	20 (10%)	4 (25%)	
Not applicable, n (%)	76 (36%)	69 (34%)		2 (1%)
Had to live rest of life with pain				
Very dissatisfied, n (%)	72 (34%)	68 (34%)	10 (63%)	
Dissatisfied, n (%)	81 (38%)	80 (40%)	1 (6%)	
Not satisfied or dissatisfied, n (%)	49 (23%)	42 (21%)	4 (25%)	
Satisfied, n (%)	12 (6%)	11 (5%)	1 (6%)	13 (6%)
Recovered from pain since it started				
Recovered, n (%)	99 (45%)	94 (47%)	7 (44%)	
Not recovered, n (%)	115 (55%)	107 (53%)	9 (56%)	13 (6%)

Table continues on next page.

TABLE 2 Baseline Descriptives (continued)

	Baseline	2 Weeks	Excluded Participants (Baseline Only)	Missing at Baseline
Previous treatment of low back pain				
Yes, n (%)	151 (71%)	145 (72%)	11 (69%)	
No, n (%)	63 (29%)	56 (28%)	5 (31%)	0 (0%)
Functional status				
RMDQ sum score, baseline, mean (SD)	9.5 (5.6)	9.6 (5.4)	11.3 (6.3)	0 (0%)
RMDQ sum score, 2 weeks, mean (SD)		6.5 (5.2)		0 (0%)
ODI sum score, baseline, mean (SD)	23.3 (16.5)	22.8 (16.5)	21.1 (16.7)	0 (0%)
ODI sum score, 2 weeks, mean (SD)		15.4 (14.7)		0 (0%)
QBPDS sum score, baseline, mean (SD)	31.3 (17.6)	31.0 (17.4)	32.7 (17.7)	0 (0%)
QBPDS sum score, 2 weeks, mean (SD)		22.1 (16.6)		0 (0%)
Comorbidities				
Heart, n (%)	24 (11%)	21 (10%)	2 (13%)	
High blood pressure, n (%)	48 (22%)	47 (23%)	3 (19%)	
Lung, n (%)	13 (6%)	13 (6%)	1 (6%)	
Diabetes, n (%)	15 (7%)	14 (7%)	4 (25%)	
Stomach, n (%)	11 (5%)	11 (5%)	2 (13%)	
Kidney, n (%)	7 (3%)	34 (17%)	3 (19%)	
Liver, n (%)	4 (1%)	4 (2%)	0 (0%)	
Blood conditions, n (%)	7 (3%)	6 (3%)	1 (6%)	
Cancer, n (%)	10 (5%)	7 (4%)	0 (0%)	
Depression, n (%)	10 (5%)	9 (5%)	0 (0%)	
Arthritis hip/knee, n (%)	49 (23%)	48 (24%)	5 (31%)	
Arthritis hand, n (%)	31 (14%)	31 (15%)	1 (6%)	
Rheumatoid arthritis, n (%)	14 (6%)	11 (55%)	1 (6%)	
Neck/shoulder problems, n (%)	96 (45%)	116 (57%)	9 (56%)	
Headache/migraine, n (%)	27 (12%)	27 (13%)	3 (19%)	
Foot problems, n (%)	48 (22%)	43 (21%)	3 (19%)	
Gout, n (%)	15 (7%)	12 (6%)	1 (6%)	
Neurological conditions (MS/Parkinson, etc), n (%)	9 (4%)	8 (4%)	0 (0%)	
Other, n (%)	17 (8%)	18 (9%)	0 (0%)	
Quality of life				
EQ-5D-5L score	0.70 (0.23)	0.71 (0.21)	0.54 (0.36)	0 (0%)
EQ-5D-5L VAS score	70.2 (16.4)	71 (16.1)	61.1 (18.2)	0 (0%)
STarT Back				
Worrying thoughts about LBP, n (%)	3.42 (2.05)	3.33 (2.0)	4.56 (2.4)	0 (0%)
Bothersomeness, moderately-extremely, n (%)	1.24 (1.34)	1.16 (1.3)	2.13 (1.7)	0 (0%)

Abbreviations: EQ-5D-5L, EuroQol 5 Dimension 5 Level; IQR, interquartile range; LBP, low back pain; MS, multiple sclerosis; n, number of participants; ODI, Oswestry Disability Index; QBPDS, Quebec Back Pain Disability Scale; RMDQ, Roland-Morris Disability Questionnaire; SD, standard deviation; VAS, visual analog scale.

The results of the test-retest reliability showed sufficient reliability as $ICC_{agreement}$ values exceeded 0.70: RMDQ, ICC(2,1) = 0.87 (95% confidence interval [CI]: 0.75, 0.94); ODI, ICC(2,1) = 0.94 (95% CI: 0.88, 0.97); QBPDS, ICC(2,1) = 0.95 (95% CI: 0.91, 0.98). The SDC (**TABLE 5**) was insufficient for all 3 instruments (RMDQ, 6.9 [30% in scale range]; ODI, 19.1 [18% in scale range]; and QBPDS, 23.6 [17% in scale range]); these are larger than previous MIC range values.[33]

Responsiveness

Responsiveness of the RMDQ, ODI, and QBPDS was tested in 193 patients. AUC values for the RMDQ of 0.75 (standard error [SE], 0.04; 95% CI: 0.68, 0.82), ODI of 0.72 (SE, 0.04; 95% CI: 0.64, 0.79), and

TABLE 3 Stable vs Non-Stable Patients

GPE:

	Participants Classified as "Stable" (n = 83)						Participants Classified as "Not Stable" (n = 110)					
	Baseline		Follow-up		Difference		Baseline		Follow-up		Difference	
Questionnaire	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
Roland-Morris Disability Questionnaire	9.9	6.1	4.8	4.9	5.1	1.2	9.2	5.3	7.8	5	1.4	0.3
Oswestry Disability Index	24.2	16.3	10.6	11.9	13.6	4.4	22.2	16.8	19	15.5	3.2	1.3
Quebec Back Pain Disability Scale	32.8	18.2	16.5	15.1	16.3	3.1	30.3	17.5	26.4	16.3	3.9	1.2

GPE:

	Participants Classified as "Stable" (n = 83)						Participants Classified as "Not Stable" (n = 109)					
	Baseline		Follow-up		Difference		Baseline		Follow-up		Difference	
Questionnaire	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
Roland-Morris Disability Questionnaire	9.9	6.1	4.7	4.9	5.2	1.2	9.1	5.3	7.8	5	1.3	0.3
Oswestry Disability Index	24.4	16.3	10.6	11.9	13.8	4.4	22.1	16.8	18.7	15.2	3.4	1.6
Quebec Back Pain Disability Scale	33	18.1	16.3	15.1	16.7	3	30.2	17.6	26.3	16.3	3.9	1.3

GPE:

	Participants Classified as "Stable" (n = 155)						Participants Classified as "Not Stable" (n = 37)					
	Baseline		Follow-up		Difference		Baseline		Follow-up		Difference	
Questionnaire	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
Roland-Morris Disability Questionnaire	9.5	5.6	6.1	5	3.4	0.6	9.5	5.9	8.4	5.3	1.1	0.6
Oswestry Disability Index	22.4	16.1	13.2	12.5	9.2	3.6	25.8	18.2	24.6	19.1	1.2	−0.9
Quebec Back Pain Disability Scale	28.3	18.3	20	15.5	8.3	2.8	32	15.5	31.5	17.5	0.5	−2

Abbreviations: GPE, Global Perceived Effect scale; SD, standard deviation.
Improved = A lot improved, Much improved; Not improved = A little improved, About the same, A little worse, Much worse, A lot worse

QBPDS of 0.75 (SE, 0.04; 95% CI: 0.67, 0.82) were obtained (**APPENDIX**).

Results of the a priori hypothesis tests on questionnaire responsiveness are presented in **TABLE 1**. All 3 PROMs confirmed the responsiveness of the instruments, meeting the 75% threshold.

A sensitivity analysis of categorizing the 7-point global change scale "a little better" answer as "Stable" instead of "Not Stable" showed an improvement in sensitivity of the SDC values (**APPENDIX**).

DISCUSSION

We evaluated the reliability, validity, and responsiveness of the RMDQ,[21] ODI,[48] and QBPDS[42] in older adults with LBP. The questionnaires have sufficient construct validity, internal consistency, test-retest reliability, and responsiveness. None of the instruments had sufficient unidimensionality for the total score following confirmatory factor analysis, with the ODI and RMDQ performing slightly better than the QBPDS. Overall, the RMDQ and ODI had superior measurement properties in older adults with LBP than the QBPDS.

TABLE 4 Confirmatory Factor Analysis

	Model	x^2	df	RMSEA	SRMR	90% CI	CFI	TLI
RMDQ	1 factor	2701.3*	230	0.057	0.121	0.048-0.067	0.934	0.927
ODI	1 factor	931.6*	45	0.068	0.0588	0.044-0.091	0.961	0.95
QBPDS	1 factor	10620.5*	170	0.153	0.088	0.174-0.192	0.883	0.869

Abbreviations: CFI, comparative fit index; CI, confidence interval; df, degrees of freedom; ODI, Oswestry Disability Index; QBPDS, Quebec Back Pain Disability Scale; RMDQ, Roland-Morris Disability Questionnaire; RMSEA, root-mean-square error of approximation; SRMR, standardized root-mean-square residual; TLI, Tucker-Lewis index.
**P<.05.*

Validity

We tested validity in 2 ways. First, we tested structural validity using confirmatory factor analysis. We found that the ODI and RMDQ had a better fit than the QBPDS.[23] Second, we tested construct validity using hypotheses that were defined a priori.[46] The construct validity of the 3 PROMs was supported by confirming 5 out of 7 (75%) of the predefined hypotheses, indicating sufficient performance.

When testing construct validity, it is important to test the construct with both

TABLE 5 Test-Retest Reliability

	Test						Retest												
	n	Min	Max	Mean	SD	SE	Cronbach α Internal Consistency (.70-.95)	n	Min	Max	Mean	SD	SE	Mean Diff	Mean Diff SD	SRM	ICC Test-Retest >0.70	ICC 95% CIs	Var	SEM	SDC	MIC	% in Scale Range
Roland-Morris Disability Questionnaire	214	0	24	9.5	5.6	0.38	.89	193	0	21	6.5	5.2	0.37	3.3	4.86	0.68	0.85	0.76	0.91	6.134	2.48	6.87	5	28.60
Oswestry Disability Index	214	0	70	23.3	16.6	1.13	.86	192	0	62	15.4	14.5	1.06	7.7	13.51	0.57	0.89	0.82	0.93	47.51	6.89	19.11	10	19.11
Quebec Back Pain Disability Scale	214	1	100	31.3	17.6	1.2	.94	193	0	79	22.1	16.6	1.19	9.4	15.94	0.59	0.84	0.76	0.9	72.37	8.51	23.58	20	23.58

Abbreviations: CIs, confidence intervals; Diff, difference; ICC, interclass correlation coefficient; Max, maximum; MIC, minimal important change; Min, minimum; n, number of participants; SD, standard deviation; SDC, smallest detectable change; SE, standard error; SEM, standard error of measurement; SRM, Standardized Response Mean; Var, Variance.

related and unrelated constructs. We tested constructs from demographics, pain, physical activity, and worrisomeness. However, we could not evaluate content validity[34] as the original design[26] of the study did not collect qualitative data. A previous systematic review highlighted the need to evaluate this property of these instruments,[5] and a head-to-head content validity study including older adults with LBP is needed. All 3 PROMs were designed for patients aged between 18 and 65 years and may not represent the physical function of the older adult. Further research on content validity in the older adult will help determine if these PROMs are appropriate for the older adult.

Reliability and Internal Consistency

The test-retest reliability results are consistent with other studies,[7,10] and values did not change with item deletion or diminish internal consistency. Due to the insufficient structural validity especially with the QBPDS, the Cronbach α should be interpreted with caution. The reliability of the Dutch RMDQ was similar to that (0.89) in a study with a longer follow-up period (9 weeks).[15] The ODI has similar reliability.[39] The QBPDS ICC value was similar to that of 0.93 reported in the original reliability study by Kopec et al.[28] The ICC value alone does not provide enough information about reliability, as the ICC is a relative reliability measure.[14] Therefore, we also calculated an absolute reliability parameter, ie, the SDC for the 3 questionnaires, which is also an estimate of measurement error. The smaller the SDC, the more free the instrument is from measurement error.[12] For instance, the SDC value of 6.9 calculated for 14 days indicates that, for a specific patient, a change of more than 7 points is most likely due to true change in the functional disability status of that patient rather than measurement error.[33]

Reliability and Measurement Error

The QBPDS, RMDQ, and ODI had excessive measurement error, as all 3 PROMs did not fall within the consensus-based MIC values.[33] Although the absolute

value of the QBPDS SDC was higher than the RMDQ and ODI, the proposed cutoff value was close to the range identified in previous work.[33] Nevertheless, MIC values of back-specific questionnaires fluctuate, depending on various features (eg, baseline scores, validity of the anchor).[11] It may be challenging for the instruments to disentangle the difference between "real" change (ie, change beyond measurement error) and "important" change, considering that the latter may fall within the range of the SDC. Therefore, we considered a change score to be minimally important only if it exceeded the SDC. SDC values can also be translated into the percentage of the scale range, and the SDC of the QBPDS would still be the smaller, equaling 17% of the 0-to-80 range, while 18% of the 0-to-100 range for the ODI and 29% of the 0-to-24 range for the RDMQ. Although similar in percentage, the QBPDS, RMDQ, and ODI displayed excessive measurement error,[7] as all 3 PROMs did not fall within the consensus-based MIC values.[33] There are no studies on the MIC of these instruments in older adults, and future studies should fill this gap.

Responsiveness

The responsiveness of the PROMs as indicated by the AUC values is similar to the sensitivity of the PROMs in other studies.[7,35] A previous systematic review on head-to-head comparison studies on the responsiveness of the ODI and RMDQ had already shown that these 2 instruments have fairly similar responsiveness.[7] On the other hand, similar head-to-head data including the QBPDS were missing.

Limitations

We used a 2-week interval to measure the test-retest reliability of the questionnaires. Although a 2-week interval is not uncommon in previous validation studies[3] of the physical function questionnaires, collecting data for a longer time frame could minimize an underestimate of the functional status scale's ability to show change (ie, lower change correlations, SRMs than studies with longer duration). A second limitation may be the limited number of comparative measures at follow-up. We collected a broad range of baseline questionnaires that were not collected at 2-week follow-up and other follow-up time points in the BACE-C study as to not overwhelm the patients with questionnaires. In hindsight, collecting a few more pain questions could have added more hypothesis data to evaluate responsiveness more comprehensively.

Although we detected no differences in participants who completed the online questionnaires as opposed to those who completed the paper version, we do not rule out the possibility that an older group of patients chose not to participate because they are limited in their computer literacy. However, the demographics of the patients included in our study are consistent with an earlier descriptive study conducted in The Netherlands.[38]

Chiropractors who participated volunteered their time and received continuing education points. Our results may reflect those practices that focus more on evidence-based practice,[25] which may compromise the generalizability of these findings. Lastly, we did not collect data on content validity, and given that all 3 questionnaires failed at some measure of validity, reliability, and responsiveness, it would be beneficial to evaluate content validity in future studies.

CONCLUSION

The RMDQ, ODI, and QBPDS have sufficient internal consistency, test-retest reliability, construct validity, and responsiveness in an older adult clinical population with LBP. The 3 instruments have similar measurement properties, but other head-to-head clinimetric studies in older adults with LBP are needed, especially in assessing content validity. ◉

KEY POINTS

FINDINGS: The Roland-Morris Disability Questionnaire (RMDQ), Oswestry Disability Index (ODI), and Quebec Back Pain Disability Scale have sufficient validity, reliability, and responsiveness in older adults with low back pain (LBP).
IMPLICATIONS: The RMDQ and ODI should be used to evaluate physical function in older adults with LBP.
CAUTION: Content validity has not yet been assessed in older adults with LBP and is a priority to confirm the validity of the 3 patient-reported outcome measures in the older adult.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: Drs Jenks, Chiarotto, Hoekstra, Ostelo, and Rubinstein were involved in study concept and design. All authors were involved in the drafting of the manuscript, critical revision of the article for important intellectual content, and final approval of the article.
DATA SHARING: Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.
PATIENT AND PUBLIC INVOLVEMENT: Patient partners were not involved in designing or conducting the study.

ACKNOWLEDGMENTS: *We would like to thank all the original authors of the BACE consortium in The Netherlands, Brazil, and Australia for the original idea for this cohort study. Additionally, we thank Jen Walraven for her comments on drafts of this manuscript.*

REFERENCES

1. Azevedo BR, Oliveira CB, Araujo GMD, et al. Is there equivalence between the electronic and paper version of the questionnaires for assessment of patients with chronic low back pain? *Spine* (Phila Pa 1976). 2020 Mar15;45(6):E329-E335. https://doi.org/10.1097/BRS.0000000000003281
2. Blikman T, Stevens M, Bulstra SK, van den Akker-Scheek I, Reininga IH. Reliability and validity of the Dutch version of the International Physical Activity Questionnaire in patients after total hip arthroplasty or total knee arthroplasty. *J Orthop Sports Phys Ther*. 2013;43:650-659. https://doi.org/10.2519/jospt.2013.4422
3. Brouwer S, Kuijer W, Dijkstra PU, Goeken LN, Groothoff JW, Geertzen JH. Reliability and stability

of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. *Disabil Rehabil*. 2004;26:162-165. https://doi.org/10.1080/09638280310001639713

4. Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine (Phila Pa 1976)*. 2011;36:S54-S68. https://doi.org/10.1097/BRS.0b013e31822ef74d
5. Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*. 2018;159:481-495. https://doi.org/10.1097/j.pain.0000000000001117
6. Chiarotto A, Deyo RA, Terwee CB, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J*. 2015;24:1127-1142. https://doi.org/10.1007/s00586-015-3892-3
7. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurment properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. *Phys Ther*. 2016;96:1620-1637. https://doi.org/10.2522/ptj.20150420
8. 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. *J Clin Epidemiol*. 2018;95:73-93. https://doi.org/10.1016/j.jclinepi.2017.11.005
9. Chiarotto A, Vanti C, Cedraschi C, et al. Responsiveness and minimal important change of the Pain Self-Efficacy Questionnaire and short forms in patients with chronic low back pain. *J Pain*. 2016;17:707-718. https://doi.org/10.1016/j.jpain.2016.02.012
10. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther*. 2002;82:8-24. https://doi.org/10.1093/ptj/82.1.8
11. de Souza IMB, Sakaguchi TF, Yuan SLK, et al. Prevalence of low back pain in the elderly population: a systematic review. *Clinics (Sao Paulo)*. 2019;74:e789. https://doi.org/10.6061/clinics/2019/e789
12. de Vet HC, Foumani M, Scholten MA, 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;68:518-524. https://doi.org/10.1016/j.jclinepi.2014.07.008
13. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes*. 2006;4:54. https://doi.org/10.1186/1477-7525-4-54
14. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Cambridge, UK: Cambridge University Press; 2011. https://doi.org/10.1017/CBO9780511996214
15. Denteneer L, Van Daele U, Truijen S, et al. The Modified Low Back Pain Disability Questionnaire: reliability, validity, and responsiveness of a Dutch language version. *Spine (Phila Pa 1976)*. 2018;43:E292-E298. https://doi.org/10.1097/BRS.0000000000002304
16. Fairbanks JC, Couper J, Davies JB, O'Brien JP. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy*. 1980;66:271-273. https://doi.org/10.1037/t04205-000
17. Franchignoni F, Giordano A, Monticone M. Construct validity of the Quebec Back Pain Disability Scale: a factor analytic and Rasch study. *Eur J Phys Rehabil Med*. 2021;57:600-606. https://doi.org/10.23736/S1973-9087.21.06502-3
18. 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:e272. https://doi.org/10.2196/jmir.9828
19. Froud R, Patel S, Rajendran D, et al. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: respice, adspice, et prospice. *PLoS One*. 2016;11:e0164573. https://doi.org/10.1371/journal.pone.0164573
20. Gabel CP, Cuesta-Vargas A, Qian M, et al. The Oswestry Disability Index, confirmatory factor analysis in a sample of 35,263 verifies a one-factor structure but practicality issues remain. *Eur Spine J*. 2017;26:2007-2013. https://doi.org/10.1007/s00586-017-5179-3
21. Gommans IHB, Koes BW, van Tulder MW. Validiteit en responsiviteit Nederlandstalige Roland Disability Questionnaire. Vragenlijst naar functionele status bij patiënten met lage rugpijn. *Ned Tijdschr Fys*. 1997;107:28-33.
22. Hicks GE, Manal TJ. Psychometric properties of commonly used low back disability questionnaires: are they useful for older adults with low back pain? *Pain Med*. 2009;10:85-94. https://doi.org/10.1111/j.1526-4637.2008.00548.x
23. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6:1-55. https://doi.org/10.1080/10705519909540118
24. Hush JM, Refshauge K, Sullivan G, De Souza L, Maher CG, McAuley JH. Recovery: what does this mean to patients with low back pain? *Arthritis Rheum*. 2009;61:124-131. https://doi.org/10.1002/art.24162
25. Innes SI, Leboeuf-Yde C, Walker BF. How comprehensively is evidence-based practice represented in councils on chiropractic education (CCE) educational standards: a systematic audit. *Chiropr Man Therap*. 2016;24:30. https://doi.org/10.1186/s12998-016-0112-0
26. Jenks AD, Hoekstra T, Axen I, et al. BAck Complaints in the Elders - Chiropractic (BACE-C): protocol of an international cohort study of older adults with low back pain seeking chiropractic care. *Chiropr Man Therap*. 2020;28:17. https://doi.org/10.1186/s12998-020-00302-z
27. Kopec JA, Esdaile JM. Functional disability scales for back pain. *Spine (Phila Pa 1976)*. 1995;20:1943-1949. https://doi.org/10.1097/00007632-199509000-00021
28. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale. Measurement properties. *Spine (Phila Pa 1976)*. 1995;20:341-352. https://doi.org/10.1097/00007632-199502000-00016
29. Kuijer W, Brouwer S, Dijkstra PU, Jorritsma W, Groothoff JW, Geertzen JH. Responsiveness of the Roland-Morris Disability Questionnaire: consequences of using different external criteria. *Clin Rehabil*. 2005;19:488-495. https://doi.org/10.1191/0269215505cr842oa
30. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006
31. Monticone M, Baiardi P, Vanti C, et al. Responsiveness of the Oswestry Disability Index and the Roland Morris Disability Questionnaire in Italian subjects with sub-acute and chronic low back pain. *Eur Spine J*. 2012;21:122-129. https://doi.org/10.1007/s00586-011-1959-3
32. Newman AN, Stratford PW, Letts L, Spadoni G. A systematic review of head-to-head comparison studies of the Roland-Morris and Oswestry measures' abilities to assess change. *Physiother Can*. 2013;65:160-166. https://doi.org/10.3138/ptc.2012-12
33. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)*. 2008;33:90-94. https://doi.org/10.1097/BRS.0b013e31815e3a10
34. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials*. 2016;17:449. https://doi.org/10.1186/s13063-016-1555-2
35. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. *Phys Ther*. 1998;78:1197-1207. https://doi.org/10.1093/ptj/78.11.1197
36. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine (Phila Pa 1976)*. 2000;25:3115-3124. https://doi.org/10.1097/00007632-200010010-00025
37. Roland M, Morris R. A study of the natural history of back pain: part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine*. 1983;8:141-144. https://doi.org/10.1097/00007632-198303000-00004

38. Rubinstein SM, Leboeuf-Yde C, Knol DL, de Koekkoek TE, Pfeifle CE, van Tulder MW. The benefits outweigh the risks for patients undergoing chiropractic care for neck pain: a prospective, multicenter, cohort study. *J Manipulative Physiol Ther.* 2007 Jul-Aug;30(6):408-18. https://doi.org/10.1016/j.jmpt.2007.04.013
39. Sandal D, Jindal R, Gupta S, Garg SK. Reliability and validity of Punjabi version of Oswestry Disability Index in patients with mechanical low back pain. *J Clin Orthop Trauma*. 2021;13:163-168. https://doi.org/10.1016/j.jcot.2020.11.011
40. Sarker K, Sethi J, Mohanty U. Effect of spinal manipulation on specific changes in segmental instability, pain sensitivity and health-related quality of life among patients with chronic non-specific low back pain – a randomized clinical trial. *Annu Res Rev Biol*. 2017;18:1-10. https://doi.org/10.9734/ARRB/2017/35926
41. Scheele J, Luijsterburg PA, Ferreira ML, et al. Back complaints in the elders (BACE); design of cohort studies in primary care: an international consortium. *BMC Musculoskelet Disord*. 2011;12:193. https://doi.org/10.1186/1471-2474-12-193
42. Schoppink LEM, van Tulder MW, Koes BW, Beurskens SAJHM, de Bie RA. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. *Phys Ther*. 1996;76:268-275. https://doi.org/10.1093/ptj/76.3.268
43. Seidel D, Brayne C, Jagger C. Limitations in physical functioning among older people as a predictor of subsequent disability in instrumental activities of daily living. *Age Ageing*. 2011;40:463-469. https://doi.org/10.1093/ageing/afr054
44. Speksnijder CM, Koppenaal T, Knottnerus JA, Spigt M, Staal JB, Terwee CB. Measurement properties of the Quebec Back Pain Disability Scale in patients with nonspecific low back pain: systematic review. *Phys Ther*. 2016;96:1816-1831. https://doi.org/10.2522/ptj.20140478
45. Stolwijk C, van Tubergen A, Ramiro S, et al. Aspects of validity of the self-administered comorbidity questionnaire in patients with ankylosing spondylitis. *Rheumatology (Oxford)*. 2014;53:1054-1064. https://doi.org/10.1093/rheumatology/ket354
46. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012
47. van der Gaag WH, Chiarotto A, Heymans MW, et al. Developing clinical prediction models for nonrecovery in older patients seeking care for back pain: the back complaints in the elders prospective cohort study. *Pain*. 2021;162:1632-1640. https://doi.org/10.1097/j.pain.0000000000002161
48. van Hooff ML, Spruit M, Fairbank JC, van Limbeek J, Jacobs WC. The Oswestry Disability Index (version 2.1a): validation of a Dutch language version. *Spine (Phila Pa 1976)*. 2015;40:E83-E90. https://doi.org/10.1097/BRS.0000000000000683
49. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health*. 2016;19:343-352. https://doi.org/10.1016/j.jval.2016.01.003
50. Wong AY, Karppinen J, Samartzis D. Low back pain in older adults: risk factors, management options and future directions. *Scoliosis Spinal Disord*. 2017;12:14. https://doi.org/10.1186/s13013-017-0121-3
51. Wong AY, Lauridsen HH, Samartzis D, Macedo L, Ferreira PH, Ferreira ML. Global consensus from clinicians regarding low back pain outcome indicators for older adults: pairwise wiki survey using crowdsourcing. *JMIR Rehabil Assist Technol*. 2019;6:e11127. https://doi.org/10.2196/11127
52. Yamato TP, Maher CG, Saragiotto BT, Catley MJ, McAuley JH. The Roland-Morris Disability Questionnaire: one or more dimensions? *Eur Spine J*. 2017;26:301-308. https://doi.org/10.1007/s00586-016-4890-9