Research Report

Measurement Properties of the Quebec Back Pain Disability Scale in Patients With Nonspecific Low Back Pain: Systematic Review

Caroline M. Speksnijder, Tjarco Koppenaal, J. André Knottnerus, Mark Spigt, J. Bart Staal, Caroline B. Terwee

C.M. Speksnijder, PT, PhD, Physical Therapy Science, Program in Clinical Health Sciences, University Medical Center Utrecht, Intern Post G05.122, PO Box 85.500, 3508 GA Utrecht, the Netherlands; Department of Oral-Maxillofacial Surgery, Prosthodontics and Special Dental Care, University Medical Center Utrecht, Utrecht, the Netherlands; and Radboud University Medical Center, IQ Healthcare, Radboud Institute for Health Sciences, Nijmegen, the Netherlands. Address all correspondence to Dr Speksnijder at: C.M.Speksnijder@umcutrecht.nl.

T. Koppenaal, PT, MSc, Department of Allied Health Professions, Fontys University of Applied Sciences, Eindhoven, the Netherlands.

J.A. Knottnerus, PhD, Department of General Practice, Care and Public Health Institute, Maastricht University, Maastricht, the Netherlands.

M. Spigt, PT, PhD, Department of General Practice, Care and Public Health Institute, Maastricht University.

J.B. Staal, PT, PhD, Research Group of Musculoskeletal Rehabilitation, HAN University of Applied Sciences, and Radboud University Medical Center, IQ Healthcare, Radboud Institute for Health Sciences.

Author information continues on next page.

Background. The Quebec Back Pain Disability Scale (QBPDS) has been translated into different languages, and several studies on its measurement properties have been done.

Purpose. The purpose of this review was to critically appraise and compare the measurement properties, when possible, of all language versions of the QBPDS by systematically reviewing the methodological quality and results of the available studies.

Method. Bibliographic databases (PubMed, Embase, CINAHL, and PsycINFO) were searched for articles with the key words "Quebec," "back," "pain," and "disability" in combination with a methodological search filter for finding studies on measurement properties concerning the development or evaluation of the measurement properties of the QBPDS in patients with nonspecific low back pain. Assessment of the methodological quality was carried out by the reviewers using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist for both the original language version of the QBPDS in English and French and all translated versions. The results of the measurement properties were rated based on criteria proposed by Terwee et al.

Results. The search strategy resulted in identification of 1,436 publications, and 27 articles were included in the systematic review. There was limited-to-moderate evidence of good reliability, validity, and responsiveness of the QBPDS for the different language versions, but for no language version was evidence available for all measurement properties.

Conclusion. For research and clinical practice, caution is advised when using the QBPDS to measure disability in patients with nonspecific low back pain. Strong evidence is lacking on all measurement properties for each language version of the QBPDS.

C.B. Terwee, PhD, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands.

[Speksnijder CM, Koppenaal T, Knottnerus JA, et al. Measurement properties of the Quebec Back Pain Disability Scale in patients with nonspecific low back pain: systematic review. *Phys Ther*. 2016;96:1816–1831.]

Published Ahead of Print:
May 26, 2016
Accepted: May 3, 2016
Submitted: October 29, 2014

One of the leading causes of disability worldwide is low back pain (LBP). Most of the time, LBP is benign and self-limiting and can be considered as nonspecific LBP, as no specific musculoskeletal pathology is found.[1–3] It occurs in similar proportions in all cultures, interferes with quality of life and work performance, and is the most common reason for medical consultation.[4,5]

To measure the construct of disability in patients with LBP, several self-report back-specific questionnaires have been developed. They are recommended by the World Health Organization as instruments to evaluate the efficacy of treatments.[4] Two of the most commonly investigated questionnaires are the Roland-Morris Disability Questionnaire (RMDQ)[6–10] and the Oswestry Low Back Pain Disability Index (ODI).[11–15] However, previous systematic reviews on available questionnaires to measure disability in patients with LBP indicate that the Quebec Back Pain Disability Scale (QBPDS)[16–18] is another well-validated and often recommended questionnaire.[17,19–21] The QBPDS also is commonly used in randomized controlled trials.[20,22–24]

The QBPDS (Appendix) was developed in 1995 in English and French.[16,17,25,26] Contrary to the RMDQ and ODI, the QBPDS is based on a conceptual model of disability.[16,17,27] The developers of the QBPDS used the World Health Organization's definition of disability as "any restriction or lack of ability to perform an activity in a manner or within the range considered normal for a human being."[28,29] Disability was operationally defined in terms of difficulty experienced while performing simple tasks.[17,30,31] During the development of the QBPDS, factor analysis of 46 items showed that the QBPDS had a 6- or 7-factor structure, with 53% of the variance explained by the first factor.[17] The decision to include 20 items in the final instrument was based on item analysis and practical considerations, which resulted in a 6-factor structure.[17] These 20 items represented 6 correlated factors, which are selected and based on the following requirements: (1) all types of physical activities relevant to back pain should be represented, including bed/rest, sitting/standing, ambulation, movement, bending/stooping, and handling large or heavy objects; and (2) the QBPDS should be highly reliable and discriminative over a wide range of disability levels, while also being practical and acceptable to both patients and clinicians.[17]

The 20 QBPDS items are scored on a 6-point scale (0="not difficult at all," 5="unable to do"). The total score is calculated by a summation of the scores for each item and ranges from 0 ("not being disabled") to 100 ("being maximally disabled").[16,17]

The QBPDS has been translated into different languages and adapted to different cultures. Studies have been performed on its measurement properties in these different adapted language versions.[32,33] A systematic review on this topic could be useful because a review on cross-cultural adaptations of the McGill Pain Questionnaire showed there is often limited evidence for the measurement properties of translated or adapted language versions. Therefore, the results from translated questionnaires should be interpreted with caution.[34,35] For the QBPDS, such a review has not yet been undertaken.

Studies of high methodological quality are needed to guarantee appropriate conclusions about measurement properties. The COSMIN checklist was developed to appraise the methodological quality of studies on measurement properties of health status questionnaires.[36,37] The purpose of this review was to critically appraise and compare the measurement properties, when possible, of the different language versions of the QBPDS for measuring disability in patients with nonspecific LBP by systematically reviewing the methodological quality and results of the available studies.

Method

Search Strategy

The following computerized bibliographic databases were searched up to September 18, 2014: PubMed (1966–2014), Embase (1974–2014), CINAHL

(EBSCOhost) (1981–2014), and PsycINFO (OvidSPhost) (1806–2014). The databases were searched with the key words "Quebec," "back," "pain," and "disability" in combination with a methodological search filter for finding studies on measurement properties (eAppendix 1, available at ptjournal.apta.org).[38] Reference lists were screened to identify additional relevant studies.

Selection Criteria

Two reviewers (T.K., M.S.) independently assessed titles, abstracts, and reference lists of the studies retrieved by the literature search. Only full-text original articles were included, primarily concerning the development or evaluation of the measurement properties of the QBPDS. Articles in all languages were included.

For inclusion, the QBPDS had to be evaluated in adult patients (≥18 years of age) with general, nonspecific LBP. Studies in patients with sciatica without any reference to a specific cause were included as well. Studies in patients with sciatica due to a specific cause (eg, nerve root comprise) or LBP due to specific causes (eg, neurological disorder, ankylosing spondylitis, fracture) were excluded. There was no minimum sample size for inclusion.

In case of disagreement between the 2 reviewers, a third reviewer (C.B.T.) made the decision regarding inclusion of the article. Both primary reviewers (T.K., M.S.) are senior physical therapists and scientists, and the third reviewer (C.B.T.) is a senior epidemiologist, which made this an optimal team for selecting articles for this review.

Quality Assessment

Assessment of the methodological quality of the included studies was carried out using the COSMIN checklist.[36,39] The COSMIN checklist consists of 9 boxes with methodological standards for how each measurement property should be assessed. Each item in a box can be scored on a 4-point scale (ie, "poor," "fair," "good," or "excellent"), which is an additional feature of the COSMIN checklist.[40] An overall score for the methodological quality of a study was determined by taking the lowest rating of any of the items in the 9 boxes. None of the studies used item response theory (IRT), so the IRT box was not used.

Table 1.
Levels of Evidence for Summary Statements on Measurement Property Based on Overall Quality[59]

Level	Rating[a]	Criteria
Strong	+++ or −−−	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or −−	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or −	One study of fair methodological quality
Conflicting	+/−	Conflicting findings
Unknown	?	Only studies of poor methodological quality

[a] +=positive rating, ?=indeterminate rating, −=negative rating.

Data extraction and assessment of (methodological) quality were independently performed by 2 reviewers (T.K. and C.B.T. for 17 of the included articles* and C.M.S. and C.B.T. for 10 of the included articles†). In case of disagreement, a third reviewer made the decision (C.M.S. for data extraction and quality assessment performed by T.K. and C.B.T. and T.K. for data extraction and quality assessment performed by C.M.S. and C.B.T.). Two reviewers (C.M.S. and C.B.T.) are senior epidemiologists and, therefore, trained in psychometrics. One reviewer (C.B.T.) is one of the developers of the COSMIN checklist, and the other reviewers (C.M.S. and T.K.) were trained by the COSMIN team on quality appraisal and data extraction.

Measurement Properties

The measurement properties are divided over 3 domains: reliability (including internal consistency, reliability, and measurement error), validity (including content validity, construct validity [ie, structural validity, hypotheses testing, and cross-cultural validity], and criterion validity), and responsiveness. Hypotheses testing was done for the original version of the QBPDS developed by Kopec and colleagues[16,17] by correlating the QBPDS with the RMDQ, ODI, Medical Outcomes Study 36-Item Short-Form survey (SF-36), and pain rated on a visual analog scale (VAS-pain), so we extracted data on the correlations of the QBPDS with these instruments for all language versions. Also related to pain, data on the correlation of the QBPDS with the Numeric Rating Pain Scale (NRPS) were included.

Part of cross-cultural validity testing concerns translation. The quality of the translation was determined by using items 4 to item 11 of the COSMIN cross-cultural validity box.

There is no gold standard for health status questionnaires available. Consequently, no level of evidence related to criterion validity can be determined for the QBPDS. The measurement properties and interpretability have been defined and discussed in detail elsewhere.[46,47] Interpretability is not a measurement property, but rather an important characteristic of a measurement instrument.[46]

Data Synthesis, Levels of Evidence, and Meta-analyses

When the quality of the translation was at least fair, we determined the quality of the measurement properties by applying levels of evidence, as defined in Table 1. The possible overall rating for a measurement property is "positive," "indeterminate," or "negative," accompanied with a level of evidence ("strong," "moderate," "limited," "conflicting," and "unknown"). To give a positive or negative rating for the results of the measurement properties, criteria for good mea-

* References 8–10, 15–18, 21, 23, 25–27, 29, 33, 41–43.
† References 5, 7, 14, 24, 30, 31, 35, 37, 44, 45.

Table 2.
Quality Criteria for Measurement Properties (Based on Terwee et al[48])[a]

Property	Rating[b]	Quality Criteria
Reliability		
Internal consistency	+	Cronbach alphas ≥.70
	?	Cronbach alpha not determined
	−	Cronbach alphas <.70
Reliability	+	ICC/weighted kappa ≥.70 OR Pearson *r*≥.80
	?	Neither ICC/weighted kappa nor Pearson *r* determined
	−	ICC/weighted kappa <.70 OR Pearson *r*<.80
Measurement error	+	MIC>SDC OR MIC outside the LOA
	?	MIC not defined
	−	MIC≤SDC OR MIC equals or inside LOA
Validity		
Content validity	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
	?	No target population involvement
	−	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	−	Factors explain <50% of the variance
Hypothesis testing	+	Correlation with an instrument measuring the same construct ≥.50 OR at least 75% of the results are in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	−	Correlation with an instrument measuring the same construct <.50 OR <75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Cross-cultural validity	+	Original factor structure confirmed OR no important DIF among language versions
	?	Confirmatory factor analysis not applied and DIF not assessed
	−	Original factor structure not confirmed OR important DIF found between language versions
Criterion validity	+	Convincing arguments that gold standard is "gold" AND correlation with gold standard ≥.70
	?	No convincing arguments that gold standard is "gold" OR doubtful design or method
	−	Correlation with gold standard <.70, despite adequate design and method
Responsiveness		
Responsiveness	+	Correlation with an instrument measuring the same construct ≥.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥0.70 AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	−	Correlation with an instrument measuring the same construct <.50 OR <75% of the results are in accordance with the hypotheses OR AUC <0.70 OR correlation with related constructs is lower than with unrelated constructs

[a] MIC=minimal important change, SDC=smallest detectable change, LOA=limits of agreement, ICC=intraclass correlation coefficient, DIF=differential item functioning, AUC=area under the curve.
[b] +=positive rating, ?=indeterminate rating, −=negative rating.

surement properties were used, based on criteria proposed by Terwee et al[48] (Tab. 2).

Meta-analyses were not performed because there were no more than 2 studies per measurement property per language version. Moreover, it is more important to evaluate whether the results are above a defined cutoff (eg, ICC>.70) than to estimate the exact pooled value of the parameter.

Results

General

The search strategy resulted in 1,436 unique publications, of which 32 articles were selected based on title and abstract. Based on the full text of these articles, 5 articles[49–53] were excluded, mainly because the articles were not about the development or evaluation of the measurement properties of the QBPDS.

Reference tracking did not result in additional articles. Finally, 27 articles[‡] were included (Fig.).

The general characteristics of these studies are presented in Table 3. One study resulted in 2 publications and, therefore, is mentioned only once in Tables 3 and 4.[16,17] Rater scores for each criterion within each COSMIN quality appraisal boxes summarized in Table 4 can be requested from the first author.

Methodological Quality

Overall, the methodological quality of the studies was fair (Tab. 4). The reviewers (T.K., C.B.T., C.M.S.) came to an agreement on all quality assessments. Only 2 studies[5,43] adequately described the measurement properties of the QPBDS. Twenty-four studies[§] did not describe how missing items were handled, and 20 studies[‖] did not describe the number or percentage of missing responses. Three studies[25,42,43] had an inadequate sample size (<50).

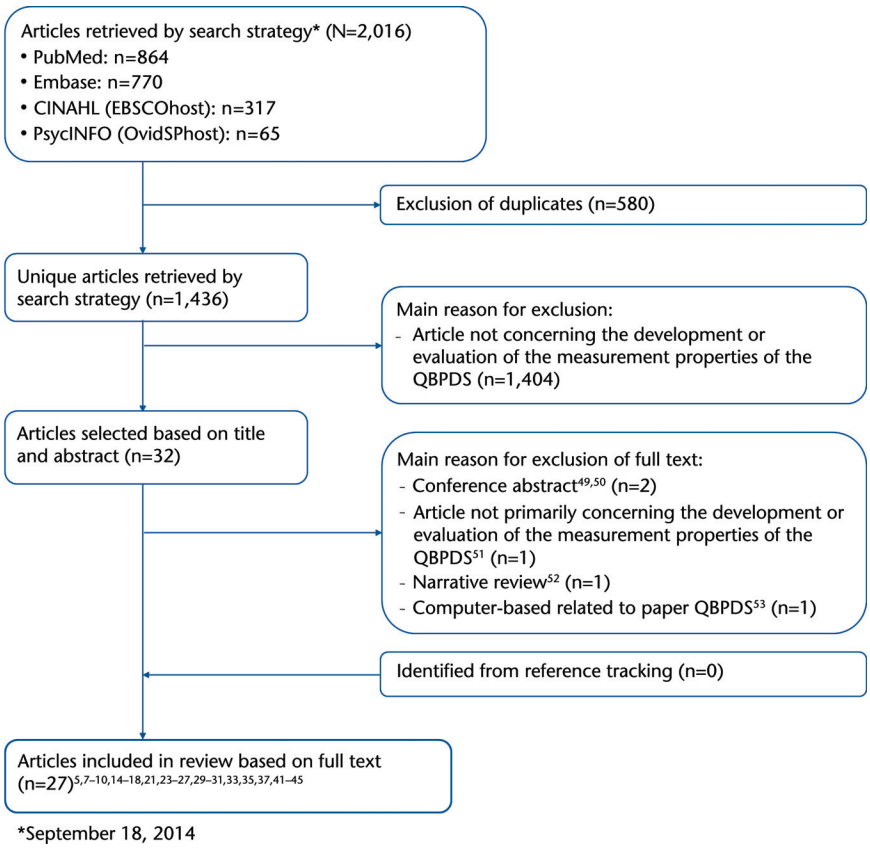

Figure.
Flowchart of search and selection. QBPDS=Quebec Back Pain Disability Questionnaire.

The QBPDS was translated into Palestinian Arabic,[44] Moroccan Arabic,[45] Chinese,[31] Dutch,[33] French,[17] Greek,[5] Hungarian,[37] Korean,[35] Persian,[25] Polish,[24] Brazilian Portuguese,[27] European Portuguese,[7] Tswana,[9] and Turkish.[14,21,41] The original English and French versions of the QBPDS developed by Kopec and colleagues[16,17] were used in 3 studies[8,15,18] and 2 studies,[42,43] respectively. The Dutch version was used in 5 studies.[10,23,26,29,33] The European Portuguese version was used in 2 studies.[7,30] All 3 studies using a Turkish version[14,21,41] used different translations. All other language versions were used in only 1 study.[#]

The results per measurement property of the QBPDS are discussed below. The results from studies of poor methodological quality (Tab. 4 and eAppendix 2, available at ptjournal.apta.org) are not mentioned because they may be biased. The data synthesis of the results and accompanying levels of evidence is presented in Table 5.

Reliability

Internal consistency. Sixteen studies[**] assessed the degree of the interrelatedness among the items of the QBPDS, expressed by Cronbach α. The studies using the English-French,[16,17] Hungarian,[37] and European Portuguese[7] language versions were of good quality, and the studies using the Greek[5] and one of the Turkish[21] language versions were of fair quality (Tab. 4). All 5 studies had positive results. The unidimensionality of the QBPDS was confirmed for the European Portuguese version by showing one predominant common factor explaining 52.1% of the variance.[7] The Cronbach α

‡ References 5, 7–10, 14–18, 21, 23–27, 29–31, 33, 35, 37, 41–45.
§ References 5, 8–10, 14–18, 21, 23–27, 29–31, 33, 35, 41–45.
‖ References 5, 8–10, 14, 15, 18, 21, 23, 24, 26, 27, 29, 30, 35, 41–45.
References 5, 9, 24, 25, 27, 31, 35, 37, 44, 45.
** References 5, 7, 9, 16, 21, 24, 25, 27, 31, 33, 35, 37, 41, 43–45.

Table 3.
Characteristics of Included Studies[a]

Study	Language	Population	Duration of Complaints $\bar{X}\pm SD$			Setting	N	Age (y) $\bar{X}\pm SD$	Male (%)	
Alnahhal et al[44]	Arabic (Gaza Strip, Palestine)	Nonspecific LBP	8.4±5.9 y			Physical therapy department of different hospitals	148	33.4±9.2	66.9	
Bendeddouche et al[45]	Arabic (Morocco)	Nonspecific chronic LBP (>3 mo)	58.4±65.9 mo			Hospital	64	47.6±12.3	37.5	
Wei et al[31]	Chinese	Nonspecific LBP (≥6 wk)	18.3±6.4 wk			Orthopedic department of a hospital	114	49.9±9.2	43.9	
Demoulin et al[10]	Dutch	Nonspecific chronic LBP (>3 mo)	24 mo (IQR: 12–72)			Rehabilitation center	223	43 (IQR: 33–49)	52.8	
Mens et al[23]	Dutch	Nonspecific lumbopelvic pain since pregnancy				Outpatient clinic of a rehabilitation center	FG: 44 SG: 56	FG: 31.7±3.2 SG: 33.5±4.9	0	
Reneman et al[26]	Dutch	Chronic nonspecific LBP	9.8±11.3 mo (range: 2–72)			Outpatient university rehabilitation center	64	38.0±8.9 (range: 23–58)	84	
Schoppink et al[33]	Dutch	Chronic LBP	0–6 mo 7–12 mo 13–24 mo >25 mo		14% 6% 8% 72%	General practice	120	39.7±10.4 (range: 21–60)	60	
van der Roer et al[29]	Dutch	Nonspecific LBP				68 primary care physical therapy practices	442	46.0±14.0	46.8	
Davidson and Keating[8]	English (Australia)	LBP		UG	IG	Physical therapy outpatient departments of 3 hospitals, 3 community health services, and 4 private physical therapy practices	UG: 47 IG: 52	UG: 55±17 (range: 19–83) IG: 49±16 (range: 20–80)	UG: 36.2 IG: 26.9	
			<1 wk	4.2%	17.3%					
			1–6 wk	21.3%	42.2%					
			6 wk–6 mo	23.4%	51.1%					
			>6 mo	19.2%	17.3%					
Fritz and Irrgang[15]	English (United States)	Acute LBP (<3 wk)	6.2±5.3 d (range: 0–19)			Physical therapy	67	39.2±9.7 (range: 21–58)	57	
Hicks and Manal[18]	English (United States)	Current LBP	31.8±23.4 wk			4 continuing care home retirement communities	107	79.6±5.7	28.1	
Kopec and colleagues[16,17,b]	English (Canada) and French (Canada)	Back pain	<1 wk 1–6 wk 6 wk–3 mo 3 mo–1 y >1 y		3.3% 16.5% 12.8% 22.7% 43.8%	Private and hospital-based physical therapy clinics; physiatry center, family group practice; orthopedic, pain, and rheumatology clinic	242	<20 20–29 30–39 40–49 50–59 60+	0.8% 18.2% 24.4% 26.0% 13.6% 16.9%	49.6

(Continued)

Table 3.
Continued

Study	Language	Population	Duration of Complaints $\bar{X}\pm SD$		Setting	N	Age (y) $\bar{X}\pm SD$	Male (%)
Wilhelm et al[42]	French (European)	Chronic LBP			Retrospective examination of medical records of outpatients with low back complaints in a physical medicine and rehabilitation department	30	41.6 (range: 29–59)	63.3
Yvanes-Thomas et al[43]	French (European)	Chronic LBP	8±5 y		Hospital pain clinic	32	4±8	66
Christakou et al[5]	Greek	Chronic LBP (>8 mo)	39.3±37.4 mo		Private rehabilitation or physical therapy clinic	130	41.1±11.6	46.2
Valasek et al[37]	Hungarian	Chronic LBP (>4 mo)	4–6 mo 7–12 mo 13–24 mo >24 mo	34.6% 24.1% 12.0% 29.3%	Outpatient clinic of the national center for spinal disorders	133	48.1±15.3	42
Suh et al[35]	Korean	Chronic LBP	21.2±9.4 mo		Hospital	80	48.8±7.5	35
Mousavi et al[25]	Persian	Chronic LBP	7.0±8.8 y (range: 0.8–40)		Physical therapy unit of a large hospital	100	40.1±11.6 (range: 17–68)	45
Misterska et al[24]	Polish	LBP due to spinal disk herniation and degenerative changes	45.9±55.5 mo		Hospital	111	41.4±11.1 (range: 21–60)	55.3
Rodrigues et al[27]	Portuguese (Brazil)	LBP	3–6 mo 6–12 mo 12–18 mo 18–24 mo >2 y	3.7% 9.2% 5.5% 11.1% 70.3%	Orthopedics and traumatology clinic	54	44.3±12.1	30
Cruz et al[7]	Portuguese (Europe)	Nonspecific chronic LBP (>3 mo)	3–6 mo 6–12 mo 12–24 mo >24 mo	13.6% 9.1% 11.4% 65.9%	16 outpatient clinics	132	46.6±12.7 (range: 18–65)	27.3
Vieira et al[30]	Portuguese (Europe)	Chronic LBP (>3 mo)	UG: ≤24 mo, n=15; 24 mo, n=29 IG: ≤24 mo, n=26; 24 mo, n=50		16 different clinical settings	UG: 41 IG: 76	UG: ≤49, n=22; >49, n=22 IG: ≤49, n=39; >49, n=37	UG: 20.5 IG: 31.6
De Beer et al[9]	Tswana	LBP			5 hospitals	100	42±9.1 (range: 23–63)	5
Bicer et al[41]	Turkish	Chronic LBP (>6 mo)	7.11±6.0 y (range: 0.50–25)		Outpatient clinics of physical medicine and rehabilitation department of a university hospital	83	43.6±10.2 (range: 18–69)	24.1

(Continued)

Table 3.
Continued

Study	Language	Population	Duration of Complaints $\bar{X}\pm SD$	Setting	N	Age (y) $\bar{X}\pm SD$	Male (%)
Düger et al[14]	Turkish	Chronic LBP		Physical therapy and rehabilitation	55	37.8±5.1	0
Melikoglu et al[21]	Turkish	LBP (>3 wk)	50.9±50.3 mo	University medical faculty physical medicine and rehabilitation department	100	45.4±15.1	26

[a] LBP=low back pain, UG=unchanged group, IG=improved group, FG=first group, SG=second group, N=included population, IQR=interquartile range.
[b] One study, evaluating the measurement properties of the Quebec Back Pain Disability Scale, resulted in 2 publications and, therefore, is mentioned once.

for the whole scale in these 5 studies ranged from .89[5] to .96.[5,16,17]

We found moderate evidence for positive internal consistency of the English, French, Hungarian, and European Portuguese language versions of the QBPDS and limited evidence for positive internal consistency of the Greek and Turkish language versions.

Reliability. Reliability (proportion of the total variance in the measurements that is due to "true" differences among patients) was evaluated in 17 studies.[††] All of these studies evaluated the test-retest reliability of the QBPDS. One of these studies also conducted an interrater reliability study, as the QBPDS was administered twice by 2 different researchers.[27] Eleven studies[‡‡] used an appropriate time interval (2–14 days) between the first and second QBPDS administrations. The reliability studies concerning the Dutch,[33] English,[18] Brazilian Portuguese,[27] and European Portuguese[7] language versions were of good quality. The studies on the Palestinian Arabic,[44] Moroccan Arabic,[45] Chinese,[31] English,[8] Greek,[5] Hungarian,[37] Persian,[25] Polish,[24] Tswana,[9] and Turkish[21] language versions were of fair quality. All of these studies[§§] had positive results and showed ICCs for test-retest reliability ranging from .70[7] to .99.[31] The study on the Brazilian Portuguese language version[27] was of good quality and evaluated both interrater and intrarater reliability. The ICCs were .96 and .93, respectively. The study on the Palestinian Arabic language version[44] was of fair quality and showed weighted kappa values of .86 and .98 for the 2 different time points in that study.

We found moderate evidence for positive reliability for the Dutch, English, Brazilian Portuguese, and European Portuguese language versions of the QBPDS and limited evidence for positive reliability for the Palestinian Arabic, Moroccan Arabic, Chinese, Greek, Hungarian, Persian, Polish, Tswana, and Turkish language versions.

Measurement error. The measurement error consists of the systematic and random error of a patient's score, which is not attributed to true changes in the construct of disability. Measurement error is calculated using data from a test-retest reliability study and is expressed in the unit of measurement of the scale (number of points on the QBPDS).[47] Measurement error was evaluated in 12 studies.[||||] Seven studies[18,27,31,33,35,37,45] used an appropriate time interval (2–14 days) between the first and second administrations of the QBPDS. The studies concerning the Dutch,[33] English,[18] and Brazilian Portuguese[27] language versions were of good quality. In the study on the English language version,[18] the smallest detectable change (SDC; $1.65 \times \sqrt{2} \times$ standard error of measurement [SEM]) was 11 points (14.5% scored below 11 points, and 0.66% scored above 89 points). In the study on the Dutch language version,[33] the limits of agreement (LOA) ranged from −15.6 to 16.4. In the study on the Brazilian Portuguese language version,[27] the intraobserver LOA ranged from −1.4 to 2.8 points, and the interobserver LOA ranged from −1.6 to 2.7 points.

In 5 fair-quality studies[8,10,29,30,37] (2 concerning the Dutch version, 1 concerning the English version, 1 concerning the Hungarian version, and 1 concerning the European Portuguese language version), SDC also was determined. In the Dutch language version study by Demoulin et al,[10] the SDC ($1.96 \times \sqrt{2} \times$ SEM) was 15.8 points. In the Dutch study by van der Roer et al,[29] the SDC ($1.96 \times \sqrt{2} \times$ SEM) was 32.9 points (95% confidence interval [CI]=24.6, 49.8) in patients with acute or subacute LBP and 24.6 points (95% CI=19.9, 32.4) in patients with chronic LBP. In the English study,[8] the SDC ($1.96 \times \sqrt{2} \times$ SEM) was 19 points (95% CI=15, 31) in patients with LBP. In the Hungarian study,[37] the SDC ($1.96 \times \sqrt{2} \times$ SEM) was 14.4 points in patients with chronic LBP. In the study on the European Portuguese language version[30]

[††] References 5, 7–9, 15–18, 21, 24, 25, 27, 31, 33, 35, 37, 44, 45.
[‡‡] References 5, 7, 16–18, 27, 31, 33, 35, 37, 44, 45.
[§§] References 5, 7–9, 18, 21, 24, 25, 27, 31, 33, 37, 44, 45.
[||||] References 8, 10, 15, 18, 27, 29–31, 33, 35, 37, 45.

Table 4.
Methodological Quality of Each Study per Measurement Property (COSMIN[a] Checklist[36])

Study	Language	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypotheses Testing	Translation	Cross-Cultural Validity	Criterion Validity	Responsiveness
Alnahhal et al[44]	Arabic (Gaza Strip, Palestine)	Poor	Fair				Poor	Poor			
Bendeddouche et al[45]	Arabic (Morocco)	Poor	Fair	Fair			Poor	Fair			
Wei et al[31]	Chinese	Poor	Fair	Fair			Fair	Fair			
Demoulin et al[10]	Dutch			Fair							Fair
Mens et al[23]	Dutch										Poor
Reneman et al[26]	Dutch						Fair				
Schoppink et al[33]	Dutch	Poor	Good	Good			Poor	Poor			Poor
van der Roer et al[29]	Dutch			Fair							
Davidson and Keating[8]	English (Australia)		Fair	Fair							Fair
Fritz and Irrgang[15]	English (United States)		Poor	Poor							Fair
Hicks and Manal[18]	English (United States)		Good	Good			Good				
Kopec and colleagues[16,17,b]	English and French (Canada)	Good	Poor		Fair	Good	Poor				Poor
Wilhelm et al[42]	French (Europe)						Poor				Poor
Yvanes-Thomas et al[43]	French (Europe)	Poor			Poor		Poor				
Christakou et al[5]	Greek	Fair	Fair		Fair	Fair	Fair	Good			
Valasek et al[37]	Hungarian	Good	Fair	Fair		Good	Fair	Fair			
Suh et al[35]	Korean	Poor	Poor	Poor			Poor	Excellent			
Mousavi et al[25]	Persian	Poor	Fair				Fair	Excellent			
Misterska et al[24]	Polish	Poor	Fair				Poor	Poor			
Rodrigues et al[27]	Portuguese (Brazil)	Poor	Good	Good			Fair	Good			
Cruz et al[7]	Portuguese (Europe)	Good	Good			Good	Good	Poor			
Vieira et al[30]	Portuguese (Europe)			Fair							Good
De Beer et al[9]	Tswana	Poor	Fair				Poor	Fair			
Bicer et al[41]	Turkish	Poor					Poor	Poor			
Düger et al[14]	Turkish						Poor				
Melikoglu et al[21]	Turkish	Fair	Fair				Fair	Excellent			

[a] COSMIN=COnsensus-based Standards for the selection of health Measurement InStruments.
[b] One study, evaluating the measurement properties of the Quebec Back Pain Disability Scale, resulted in 2 publications and, therefore, is mentioned once.

Table 5.
Data Synthesis[48]: Levels of Evidence of Overall Quality of the QBPDS Measurement Properties per Language[a]

Language	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypothesis Testing				Translation	Cross-Cultural Validity	Criterion Validity	Responsiveness
						RMDQ	ODI	SF-36	VAS-Pain/ NRPS				
Arabic (Gaza Strip, Palestine)[44]	?	+	na	na	na	na	?	na	?	?	na	na	na
Arabic (Morocco)[45]	?	+	?*	na	na	?	na	na	?	+	na	na	na
Chinese[31]	?	+	?*	na	na	na	+	na	+	+	na	na	na
Dutch[10,23,26,29,33]	?	++	−−	na	na	+	+	na	?	?	na	na	+
English[8,15–18]	++	++	?*	+	++	?	?	++	?	na	na	na	+
French[16,17,42,43]	++	?	na	+	++	?	?	?	?	na	na	na	?
Greek[5]	+	+	na	+	+	+	na	na	na	++	na	na	na
Hungarian[37]	++	+	?*	na	++	na	+	na	+	+	na	na	na
Korean[35]	?	?	?	na	na	na	?	?	?	+++	na	na	na
Persian[25]	?	+	na	na	na	+	+	+	−	+++	na	na	na
Polish[24]	?	+	na	na	na	na	?	na	na	?	na	na	na
Portuguese (Brazil)[27]	?	++	?*	na	na	+	na	na	+	++	na	na	na
Portuguese (Europe)[7,30]	++	++	−	na	++	++	na	na	−−	?	na	na	++
Tswana[9]	?	+	na	na	na	na	na	na	?	+	na	na	na
Turkish[41]	?	na	na	na	na	na	na	na	?	?	na	na	na
Turkish[14]	na	na	na	na	na	na	?	na	?	na	na	na	na
Turkish[21]	+	+	na	na	na	na	+	na	−	+++	na	na	na

[a] +++ or −−−=strong evidence of positive/negative result, ++ or −−=moderate evidence of positive/negative result, + or −=limited evidence of positive/negative result, +/−=conflicting evidence, ?=result indeterminate or unknown due to poor methodological quality, ?*=result unknown due to lacking information about the minimal important change, na=no information available, QBPDS=Quebec Back Pain Disability Questionnaire, RMDQ=Roland-Morris Disability Questionnaire, ODI=Oswestry Low Back Pain Disability Index, VAS=visual analog scale, NRPS=Numeric Rating Pain Scale.

in patients with acute or subacute LBP, the SDC ($1.65 \times \sqrt{2} \times$ SEM) was 19 points. In 2 other fair studies (Moroccan Arabic[45] and Chinese[31]), LOA also were reported. The intraobserver LOA of the Moroccan Arabic language version[45] ranged from −19.3 to 20.7 points. The intraobserver LOA of the Chinese language version[31] ranged from −17.1 to 18.1 points.

We found moderate evidence for a negative measurement error for the Dutch language version. However, for the Moroccan Arabic, Chinese, English, Hungarian, and Brazilian Portuguese language versions of the QBPDS, we could not perform a best evidence synthesis because of the lack of a minimal important change (MIC).

Validity

Content validity (including face validity). The aim of the study by Kopec et al[17] was to develop a new scale of functional disability associated with back pain. All 20 items of the QBPDS reflect disabilities in performing activities in LBP well (Appendix). The studies on the Palestinian Arabic[44] and Greek[5] language versions concluded that the QBPDS assesses the intended construct. In particular, patients agreed that the scale seemed to be a reasonable test for evaluating the functional disability of patients with LBP. In all other studies, it was not reported whether patients agreed that the scale appeared to be a reasonable test for evaluating the functional disability of patients with LBP.

Three studies[5,17,43] evaluated the degree to which the content of the QBPDS is an adequate reflection of construct disability. Two studies (English-French[16,17] and Greek[5]) were of at least fair quality. From the study on the development of the QBPDS[16,17] and the study using the Greek version[5] of the QBPDS, it can be deduced that all items were considered relevant to measure the construct disability in a population of patients with LBP, as rated by experts and patients, and no important items were missing. We found limited evidence for positive content validity for the English, French, and Greek language versions of the QBPDS.

Construct validity (including structural validity, hypothesis testing, and cross-cultural validity). To assess structural validity, 3 methodologically good studies (English-French,[16,17] Hungarian,[37] European Portuguese[7]) and 1 fair study (Greek[5]) assessed whether the scores of the QBPDS are an adequate reflection of the dimensionality of construct disability. The studies on the English-French[16,17] and the European Portuguese[7] language versions suggested that the 20-item scale could be considered approximately unidimensional. The study on the Hungarian language version[37] showed a 4-factor structure of the 20 items. The studies on the original English-French version[16,17] and the Greek language version[5] showed a 6-factor structure of the 20 items.

We found moderate evidence for positive structural validity for the English, French, Hungarian, and European Portuguese language versions of the QBPDS and limited evidence for positive structural validity for the Greek language version of the QBPDS.

Twenty studies[##] tested hypotheses of the relation between the QBPDS and other measurement tools. However, in 17 studies,[***] hypotheses were vaguely or not described. Only 12 of the 20 studies[†††] adequately described the construct of the comparator instrument. Two studies (English[18] and European Portuguese[7]) were of good quality, and 7 studies (Chinese,[31] Dutch,[26] Greek,[5] Hungarian,[37] Persian,[25] Brazilian Portuguese,[27] and Turkish[21]) were of fair quality. In the studies on the Dutch,[26] Greek,[5] Persian,[25] Brazilian Portuguese,[27] and European Portuguese[7] versions, the relation between the QBPDS and the RMDQ was tested. The correlations between the QBPDS and RMDQ ranged from .60[25,26] to .85,[27] as supposed. The relation between the QBPDS and the ODI was tested in the studies using the Chinese,[31] Dutch,[26] Hungarian,[37] Persian,[25] and Turkish[21] language versions. The correlations between the QBPDS and ODI ranged, as supposed, from .67[21] to .90.[31] The studies using the English[18] and Persian language versions[25] showed correlations between the QBPDS and SF-36 ranging from .64[18] to .69,[25] as we expected. The studies using the Chinese,[31] Hungarian,[37] Brazilian Portuguese,[27] European Portuguese,[7] and Turkish[21] language versions showed correlations between the QBPDS and VAS-pain ranging from .37[21] to .87.[31]

Correlations with the VAS-pain were expected to be lower than the correlations with disability measures, and indeed the correlation with the VAS-pain was .37 in one study on the Turkish language version[21] and .38 in the study on the European Portuguese language version.[7] However, in the studies on the Chinese,[31] Hungarian,[37] and Brazilian Portuguese[27] language versions, the correlations of the QBPDS with the VAS-pain were higher than expected (.62,[37] .75,[27] and .87[31]). Also, the correlation of the QBPDS with the bodily pain subscale of the SF-36 was .50 in the study on the English language version[18] and .62 in the study on the Persian language version.[25] As expected, low correlations were found between the QBPDS and the SF-36 mental health (.25[18] and .40[25]) and role-emotional functioning (.26[18] and .37[25]) subscales.

We found moderate evidence for positive construct validity for the English language version related to the SF-36. For the Tswana language version, we found moderate evidence for positive construct validity related to the RMDQ and negative construct validity related to the VAS-pain. Limited evidence was found for positive construct validity for the Chinese (ODI, VAS-pain), Dutch (RMDQ, ODI), Greek (RMDQ), Hungarian (ODI, VAS-pain), Persian (RMDQ, ODI, SF-36), Portuguese Brazilian (RMDQ, VAS-pain), and Turkish (ODI) language versions. However, we also found limited evidence for negative construct validity for the Persian and Turkish language versions related to VAS-pain.

For cross-cultural validity and translation, none of the included studies assessed whether the performance of the items

[##] References 5, 7, 9, 14, 16, 18, 21, 24–27, 31, 33, 35, 37, 41–45.
[***] References 5, 7, 9, 14, 18, 21, 24–27, 31, 33, 35, 41–44.
[†††] References 5, 7, 14, 18, 21, 24–26, 33, 41–43.

on a translated QBPDS were an adequate reflection of the performance of the original QBPDS (cross-cultural validity) (eg, by using multiple group factor analyses of evaluating differential item functioning).

In 15 studies,[‡‡‡] a translation of the QBPDS was described. The QBPDS was translated into Palestinian Arabic,[44] Moroccan Arabic,[45] Chinese,[31] Dutch,[33] French,[17] Greek,[5] Hungarian,[37] Korean,[5] Brazilian Portuguese,[27] European Portuguese,[7] Persian,[25] Polish,[24] Tswana,[9] and Turkish.[14,21,41] The Korean,[35] Persian,[25] and Turkish[21] translation studies were of excellent quality; the Brazilian Portuguese[27] and Greek[5] translation studies were of good methodological quality; and the Moroccan Arabic,[45] Chinese,[31] Hungarian,[37] and Tswana[9] translation studies were of fair methodological quality. We found no evidence for cross-cultural validity.

Criterion validity. As stated in the Method section, no gold standard for health status questionnaires is available.

Responsiveness

Eight studies[8,10,15–17,23,30,33,42] evaluated the ability of the QBPDS to detect change over time in the construct of disability. One study on the European Portuguese language version[30] was of good quality and showed an area under the receiver operating characteristic (ROC) curve (AUC) of 0.74. In the European Portuguese language version,[30] the AUC was interpreted as the probability of correctly discriminating between "clinically stable" (score≤4) and "clinically improved" (score≥5) patient outcomes, based on the change in scores on the Patient Global Improvement Change Scale (PGIC-PT; ordinal scale from 1 to 7[54]). One study on the Dutch language version[10] and 2 studies on the English language version[8,15] were of fair quality and showed positive results, with AUCs of .85,[10] .74,[8] and .87.[15] In the Dutch language version,[10] the AUC was interpreted as the probability of correctly discriminating between "clinically stable" and "clinically improved" patient outcomes, using a change score (score≥6) and an unchanged score (score=3–5) of the following ordinal scale: 1="worse than ever," 2="much worsened," 3="slightly worsened," 4="unchanged," 5="slightly improved," 6="much improved," and 7="completely recovered." In the English language version of the QBPDS in the study by Davidson and Keating,[8] the AUC was interpreted as the probability of correctly discriminating between "unchanged" and "improved" patient outcome, using a change score (score≤3) and an unchanged score (score=4–6) of the following ordinal scale: 1="completely gone," 2="much better," 3="better," 4="a little better," 5="about the same," 6="a little worse," and 7="much worse." In the English language version of the QBPDS in the study by Fritz and Irrgang.[15] The AUC was interpreted as the probability of correctly discriminating between "clinically stable" and "clinically improved" patient outcomes based on the 15-point rating scale of Jaeschke et al,[55] using a change score (score≥3) and an unchanged score (score=−3 to 3) of the following ordinal scale: 1="completely gone," 2="much better," 3="better," 4="a little better," 5="about the same," 6="a little worse," and 7="much worse."

In 3 of these studies, patients were treated by physical therapists for LBP,[8] acute LBP,[15] and chronic LBP.[30] In one study,[10] patients were treated by a multidisciplinary team for chronic LBP.

We found moderate evidence for positive responsiveness for the European Portuguese language version and limited evidence for positive responsiveness for the Dutch and English language versions of the QBPDS.

Interpretability: MIC

In one study on the European Portuguese language version[30] and 2 studies on the Dutch language version,[10,29] the MIC of the QBPDS was estimated. These 3 studies used the ROC method to determine the MIC. In the European Portuguese study by Vieira et al[30] and the Dutch study by Demoulin et al,[10] the MIC was determined by identifying the point closest to the upper left corner on the ROC curve. In the Dutch study by van der Roer et al,[29] the MIC was determined by the optimal cutoff point as that point that yields the lowest overall misclassification.

In the European Portuguese language version of the QBPDS,[30] the AUC was interpreted as the probability of correctly discriminating between "clinically stable" (score≤4) and "clinically improved" (score≥5) patient outcomes, based on the change in the PGIC-PT (ordinal scale from 1 to 7[54]) score. In the Dutch study by Demoulin et al,[10] the AUC was interpreted as the probability of correctly discriminating between "clinically stable" and "clinically improvement" patient's outcome, using a change score (score≥6) and an unchanged score (score=3–5) of the following ordinal scale: 1="worse than ever," 2="much worsened," 3="slightly worsened," 4="unchanged," 5="slightly improved," 6="much improved," and 7="completely recovered." In the Dutch study by van der Roer et al,[29] the AUC was interpreted as the probability of correctly discriminating between "stable" and "improved" patient outcomes, using a change score (score ≤2) and an unchanged score (score=3–5) of the following ordinal scale: 1="completely recovered," 2="much improved," 3="slightly improved," 4="no change," 5="slightly worsened," 6="much worse."

In the study on the European Portuguese version,[30] the MIC was defined as 6.5 points (AUC=0.74) for patients with chronic LBP after 6 weeks. In one Dutch study,[10] the MIC was defined as 5 points (AUC=0.85) or an 18.1% change from baseline (AUC=0.86) after 10 weeks in patients with chronic LBP who received multidisciplinary rehabilitation. In the other Dutch study,[29] the MIC was defined as 17.5 points (AUC=0.74) for patients with acute or subacute LBP and 8.5 points for patients with chronic LBP after 12 weeks.[29] Patients of the 2 last-mentioned studies received physical therapy. An expert panel recommended an MIC of 20 points or a change of 30% from baseline.[22]

In 3 of these studies, patients were treated by physical therapists for LBP,[8]

[‡‡‡] References 5, 7, 9, 17, 21, 24, 25, 27, 31, 33, 35, 37, 41, 44, 45.

acute LBP,[15] and chronic LBP.[30] In one study,[10] patients were treated by a multidisciplinary team for chronic LBP.

Discussion

There is limited-to-moderate evidence for good reliability, validity, and responsiveness of the QBPDS for different language versions. However, there is no complete evidence for all measurement properties in any language version of the QBPDS. Because of the wide ranges in SDC and MIC values, it is difficult to determine whether the QBPDS can distinguish true changes from the systematic and random error of a score in individual patients.

Concerning the degree to which the QBPDS measures the construct of disability,[7,18] the construct of the QBPDS also seems to be correlated to bodily pain.[18] By using the term "difficulty," possibly both the constructs disability and pain are measured by the QBPDS.

Limitations

A limitation of this study is that we were not able to differentiate among study settings, follow-up durations, interventions, and subacute, acute, and chronic LBP. There were not enough studies to enable distinctions among these groups. Therefore, we are not sure that the same results apply to, for example, patients with acute and chronic LBP. One study of poor quality (because of a small sample size) measuring patients with acute LBP showed an ICC of .55.[15] This finding may suggest that the reliability of the QBPDS is not as good in patients with acute LBP, but more evidence is needed in good-quality studies.

For this systematic review, we were interested in all language versions; however, we used only English terms in our search string to identify relevant articles, which could have limited the inclusion of non-English studies.

The QBPDS was translated into 14 different languages.[§§§] The translation of the original version was of at least good quality in 5 studies.[5,21,25,27,35] However, cross-cultural validity has not been assessed. It is therefore unknown if the translated versions of the QBPDS assesses disability in the same manner as its original version. Cross-cultural validity can be assessed by determining if the factor structure of the translated version equals the original factor structure (in a multiple-group factor analysis), or by assessing if there is differential item functioning between the 2 versions.[47] Differential item functioning means that patients with the same true score on the construct have the same score on the measurement instrument item.[39]

We recommend these statistical analyses for each language version to show if the scores of the translated QBPDS versions can be interpreted in the same way as the original version of the QBPDS developed by Kopec and colleagues.[16,17]

The decision to include 20 items in the final instrument was based on item analysis and practical considerations, which resulted in a 6-factor structure.[17] Two good-quality studies, one on the original English-French version[16,17] and one on the European Portuguese version,[7] suggested that the 20-item scale could be considered approximately unidimensional, explaining 52% to 53% of the variance. However, another good-quality study using the Hungarian version[37] showed a 4-factor structure (everyday activities, ambulation, sitting/carrying, and bed/rest) of the 20 QBPDS items, and the fair-quality studies on the original English-French version[17] and the Greek language version[5] showed a 6-factor structure (movement, handling of large/heavy objects, bending/stooping, ambulation, sit/stand, and bed/rest) of the 20 QBPDS items. An explanation for these different results may be low cross-cultural validity. As the dimensional structure of the QBPDS, therefore, is not entirely clear, the results on internal consistency should be interpreted with caution because unidimensionality is a prerequisite for a clear interpretation of the internal consistency statistics.[56]

Future Research

For almost all measurement properties, additional studies are needed. Foremost, studies are needed to determine cross-cultural validity so that scores related to different language versions of the QBPDS can be compared with each other.

We recommend adequate factor analyses for each language version in future studies of the QBPDS. Also, IRT analyses are recommended to investigate the internal structure of the QBPDS.

Studies that determine the reliability to determine "true" differences between patients and measurement error to distinguish true changes from systematic and random error also are needed.[46,57] Regarding interpretability, only 3 studies[10,29,30] of at least fair quality determined the MIC of the QBPDS, and their results varied widely. Because of the wide range in SDC and MIC values, it is difficult to conclude whether the SDC is larger or smaller than the MIC. It is recommended, therefore, that future research should focus on determining the MIC and SDC in all language versions. Until there is more evidence regarding the MIC, we recommend using the conservative guidelines as recommended by the expert panel (an MIC of 20 points or a change of 30% from baseline).[22]

Furthermore, more high-quality studies on responsiveness are needed. Most studies assessing responsiveness of the QBPDS did not formulate or only vaguely formulated hypotheses regarding expected correlations in advance.[8,16,33,42] Without specific hypotheses, the risk of bias is high because retrospectively it is tempting to come up with alternative explanations for low correlations instead of concluding that the questionnaire is not responsive.[48] We, therefore, recommend performing additional high-quality studies, testing specific hypotheses regarding the correlation of changes in the QBPDS with changes in other disability questionnaires, pain measures, and measures of psychosocial functioning.

The original version of the QBPDS was in English and French. For these language versions, research of at least good methodological quality is needed for every measurement property mentioned by the COSMIN checklist.[36] Internal consistency, reliability, measurement error,

[§§§] References 5, 7, 9, 17, 21, 24, 25, 27, 31, 33, 35, 37, 41, 44, 45.

structural validity, and hypotheses testing have to be investigated once again by a study of at least good quality. Thereby, content validity, criterion validity, responsiveness, and interpretability have to be investigated once by a study of excellent methodological quality or at least twice by good methodological studies.

It is advisable to perform similar reviews for the RMDQ and the ODI to enhance the comparability among the different questionnaires. Finally, future research should focus on comparing the QBPDS, RMDQ, and ODI with newly developed, IRT-based instruments, such as the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Functioning instruments (http://www.nihpromis.org). The PROMIS offers major advantages, such as the possibility of computer adaptive testing and comparability of scores across patient populations. Research has shown that the PROMIS has better measurement properties than traditional questionnaires.[58] Although the PROMIS is a generic questionnaire, it may be as responsive as disease-specific questionnaires when used as a computerized adaptive test, and its responsiveness should be investigated in future studies.

For research and clinical practice, we advise using the QBPDS with caution to measure disability in patients with nonspecific LBP. Strong evidence is lacking on all measurement properties for each language version of the QBPDS.

Dr Speksnijder, Mr Koppenaal, Professor Knottnerus, Dr Spigt, and Dr Terwee provided concept/idea/research design. All authors provided writing and data analysis. Dr Speksnijder, Mr Koppenaal, Dr Spigt, and Dr Terwee provided data collection. Dr Speksnijder and Mr Koppenaal provided project management. Professor Knottnerus provided facilities/equipment and institutional liaisons. Mr Koppenaal, Professor Knottnerus, and Dr Terwee provided consultation (including review of manuscript before submission).

The authors thank Alice Tillema (Medical Library; Radboudumc Nijmegen, the Netherlands) for optimizing the search filter.

DOI: 10.2522/ptj.20140478

References

1 Ozguler A, Leclerc A, Landre MF, et al. Individual and occupational determinants of low back pain according to various definitions of low back pain. *J Epidemiol Community Health*. 2000;54:215–220.

2 Griffith LE, Hogg-Johnson S, Cole DC, et al. Low-back pain definitions in occupational studies were categorized for a meta-analysis using Delphi consensus methods. *J Clin Epidemiol*. 2007;60:625–633.

3 Urquhart DM, Hoving JL, Assendelft WW, et al. Antidepressants for non-specific low back pain. *Cochrane Database Syst Rev*. 2008;1:CD001703.

4 Ehrlich GE. Low back pain. *Bull World Health Organ*. 2003;81:671–676.

5 Christakou A, Andriopoulou M, Asimakopoulos P. Validity and reliability of the Greek version of the Quebec Back Pain Disability Scale. *J Back Musculoskelet Rehabil*. 2011;24:145–154.

6 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 (Phila Pa 1976)*. 1983;8:141–144.

7 Cruz EB, Fernandes R, Carnide F, et al. Cross-cultural adaptation and validation of the Quebec Back Pain Disability Scale to European Portuguese language. *Spine (Phila Pa 1976)*. 2013;38:E1491–E1497.

8 Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther*. 2002;82:8–24.

9 De Beer N, Stewart A, Becker P. Validation of the Tswana versions of the Roland-Morris Disability Questionnaire, Quebec Disability Scale and Waddell Disability Index. *S Afr J Physiother*. 2008;64:23–30.

10 Demoulin C, Ostelo R, Knottnerus JA, Smeets RJ. Quebec Back Pain Disability Scale was responsive and showed reasonable interpretability after a multidisciplinary treatment. *J Clin Epidemiol*. 2010;63:1249–1255.

11 Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66:271–273.

12 Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;25:2940–2952.

13 Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research: a proposal for standardized use [erratum in: *Spine (Phila Pa 1976)*. 1999;24:418]. *Spine (Phila Pa 1976)*. 1998;23:2003–2013.

14 Düger T, Yilmax O, Aki E, et al. Two different back pain scales and their relation to pain intensity in women. *Pain Clinic*. 2005;17:117–120.

15 Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale [erratum in: *Phys Ther*. 2008;88:138–139]. *Phys Ther*. 2001;81:776–788.

16 Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale: measurement properties. *Spine (Phila Pa 1976)*. 1995;20:341–352.

17 Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale: conceptualization and development. *J Clin Epidemiol*. 1996;49:151–161.

18 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.

19 Grotle M, Brox JI, Vollestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine (Phila Pa 1976)*. 2005;30:130–140.

20 Rocchi MB, Sisti D, Benedetti P, et al. Critical comparison of nine different self-administered questionnaires for the evaluation of disability caused by low back pain. *Eura Medicophys*. 2005;41:275–281.

21 Melikoglu MA, Kocabas H, Sezer I, et al. Validation of the Turkish version of the Quebec Back Pain Disability Scale for patients with low back pain. *Spine (Phila Pa 1976)*. 2009;34:E219–E224.

22 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.

23 Mens JM, Vleeming A, Snijders CJ, et al. Responsiveness of outcome measurements in rehabilitation of patients with posterior pelvic pain since pregnancy. *Spine (Phila Pa 1976)*. 2002;27:1110–1115.

24 Misterska E, Jankowski R, Glowacki M. Quebec Back Pain Disability Scale, Low Back Outcome Score and revised Oswestry Low Back Pain Disability Scale for patients with low back pain due to degenerative disc disease. *Spine (Phila Pa 1976)*. 2011;36:E1722–E1729.

25 Mousavi SJ, Parnianpour M, Mehdian H, et al. The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions. *Spine (Phila Pa 1976)*. 2006;31:E454–E459.

26 Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. *J Occup Rehabil*. 2002;12:119–129.

27 Rodrigues MF, Michel-Crosato E, Cardoso JR, Traebert J. Psychometric properties and cross-cultural adaptation of the Brazilian Quebec Back Pain Disability Scale Questionnaire. *Spine (Phila Pa 1976)*. 2009;34:E459–E464.

28 Wood PH. Appreciating the consequences of disease: the *International Classification of Impairments, Disabilities, and Handicaps*. *WHO Chron*. 1980;34:376–380.

29 van der Roer N, Ostelo RW, Bekkering GE, et al. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine (Phila Pa 1976)*. 2006;31:578–582.

30 Vieira AC, Moniz S, Fernandes R, et al. Responsiveness and interpretability of the Portuguese version of the Quebec Back Pain Disability Scale in patients with chronic low back pain. *Spine (Phila Pa 1976)*. 2014;39:E346–E352.

31 Wei X, Yi H, Wu B, et al. A valid cross-culturally adapted simplified Chinese version of the Quebec Back Pain Disability Scale. *J Clin Epidemiol*. 2012;65:1321–1328.

32 Costa LO, Maher CG, Latimer J. Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine (Phila Pa 1976)*. 2007;32:1028–1037.

33 Schoppink LE, vanTulder MW, Koes BW, et al. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. *Phys Ther*. 1996;76:268–275.

34 Menezes Costa LC, Maher CG, McAuley JH, Costa LO. Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. *J Clin Epidemiol*. 2009;62:934–943.

35 Suh KT, Kim JI, Lim JM, et al. Validation of the Korean version of the Quebec Back Pain Disability Scale. *J Spinal Disord Tech*. 2012;25:447–450.

36 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539–549.

37 Valasek T, Varga PP, Szoverfi Z, et al. Reliability and validity study on the Hungarian versions of the Oswestry Disability Index and the Quebec Back Pain Disability Scale. *Eur Spine J*. 2013;22:1010–1018.

38 Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18:1115–1123.

39 de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine; Practical Guides to Biostatistics and Epidemiology*. New York, NY: Cambridge University Press; 2011.

40 Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21:651–657.

41 Bicer A, Yazici A, Camdeviren H, et al. Assessment of pain and disability in patients with chronic low back pain: reliability and construct validity of the Turkish version of the Quebec Back Pain Disability Scale and Pain Disability Index. *J Back Musculoskelet Rehabil*. 2005;18:37–44.

42 Wilhelm F, Fayolle-Minon I, Phaner V, et al. Sensitivity to change of the Quebec Back Pain Disability Scale and the Dallas Pain Questionnaire. *Ann Phys Rehabil Med*. 2010;53:15–23.

43 Yvanes-Thomas M, Calmels P, Bethoux F, et al. Validity of the French-language version of the Quebec Back Pain Disability Scale in low back pain patients in France. *Joint Bone Spine*. 2002;69:397–405.

44 Alnahhal A, May S. Validation of the Arabic Version of the Quebec Back Pain Disability Scale. *Spine (Phila Pa 1976)*. 2012;37:E1645–E1650.

45 Bendeddouche I, Rostom S, Bahiri R, et al. Translation, adaptation and validation of the Moroccan version of the Quebec Back Pain Disability Scale. *Clin Rheumatol*. 2012;31:943–949.

46 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.

47 Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22.

48 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.

49 Lazary A, Boja S, Szilagyi A, et al. Efficacy of a school-based early prevention program on spinal functional capicity. *Eur Spine J*. 2012;21:S269–S337.

50 Bendeddouche I, Rostom S, Bahiri R, et al. Translation, adaptation and validation of the Moroccan version of the Quebec Back Pain Disability Scale. *Ann Phys Rehabil Med*. 2011;54(suppl 1):e255–e269.

51 Sanchez K, Papelard A, Nguyen C, et al. Patient-preference disability assessment for disabling chronic low back pain: a cross-sectional survey. *Spine (Phila Pa 1976)*. 2009;34:1052–1059.

52 Smeets R, Koke A, Lin CW, et al. Measures of function in low back pain/disorders: low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive Isoinertial Lifting Evaluation (PILE), Quebec Back Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire (RDQ). *Arthritis Care Res (Hoboken)*. 2011;63:S158–S173.

53 Bent H, Ratzlaff CR, Goligher EC, et al. Computer-administered bath ankylosing spondylitis and Quebec Scale outcome questionnaires for low back pain: agreement with traditional paper format. *J Rheumatol*. 2005;32:669–672.

54 Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. *J Manipulative Physiol Ther*. 2004;27:26–35.

55 Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415.

56 Cortina JM. What is coefficient alpha: an examination of theory and applications. *J Appl Psychol*. 1993;78:98–104.

57 de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033–1039.

58 Fries JF, Krishnan E, Rose M, et al. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Res Ther*. 2011;13:R147.

59 van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)*. 2003;28:1290–1299.

Appendix.
Quebec Back Pain Disability Scale (Kopec et al[16,17])[a]

This questionnaire is about the way your back pain is affecting your daily life. People with back problems may find it difficult to perform some of their daily activities. We would like to know if you find it difficult to perform any of the activities listed below, because of your back. Please choose one response option for each activity (do not skip any activities) and mark the corresponding box with an X.

Today, would you find it difficult to perform the following activities because of your back?

	Not difficult at all	Minimally difficult	Somewhat difficult	Fairly difficult	Very difficult	Unable to do
1.Get out of bed	☐	☐	☐	☐	☐	☐
2.Sleep through the night	☐	☐	☐	☐	☐	☐
3.Turn over in bed	☐	☐	☐	☐	☐	☐
4.Ride in a car	☐	☐	☐	☐	☐	☐
5.Stand up for 20-30 minutes	☐	☐	☐	☐	☐	☐
6.Sit in a chair for several hours	☐	☐	☐	☐	☐	☐
7.Climb one flight of stairs	☐	☐	☐	☐	☐	☐
8.Walk a few blocks (300-400 m)	☐	☐	☐	☐	☐	☐
9.Walk several miles	☐	☐	☐	☐	☐	☐
10.Reach up to high shelves	☐	☐	☐	☐	☐	☐
11.Throw a ball	☐	☐	☐	☐	☐	☐
12.Run one block (about 100 m)	☐	☐	☐	☐	☐	☐
13.Take food out of the refrigerator	☐	☐	☐	☐	☐	☐
14.Make your bed	☐	☐	☐	☐	☐	☐
15.Put on socks (pantyhose)	☐	☐	☐	☐	☐	☐
16.Bend over to clean the bathtub	☐	☐	☐	☐	☐	☐
17.Move a chair	☐	☐	☐	☐	☐	☐
18.Pull or push heavy doors	☐	☐	☐	☐	☐	☐
19.Carry two bags of groceries	☐	☐	☐	☐	☐	☐
20.Lift and carry a heavy suitcase	☐	☐	☐	☐	☐	☐

[a] QBPDS © Jacek A. Kopec, 1995. All rights reserved. Used with permission from © Mapi Research Trust, Lyon, France: https://eprovide.mapi-trust.org.