

## ORIGINAL REPORT

# DO PATIENT-REPORTED OUTCOME MEASURES DESCRIBE FUNCTIONING IN PATIENTS WITH LOW BACK PAIN, USING THE BRIEF INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY AND HEALTH CORE SET AS A REFERENCE?

Charlotte Ibsen, PT, MSc<sup>1,2</sup>, Berit Schiøttz-Christensen, MD, PhD<sup>3,4</sup>, Hanne Melchiorsen, RN, MPH<sup>2</sup>, Claus Vinther Nielsen, MD, PhD<sup>1,2</sup> and Thomas Maribo, MSc, PhD<sup>1,2</sup>

From the <sup>1</sup>Department of Public Health, Section of Clinical Social Medicine and Rehabilitation, Aarhus University, <sup>2</sup>DEFACTUM, Central Denmark Region, <sup>3</sup>Spine Centre of Southern Denmark, Hospital Lillebaelt, Middelfart and <sup>4</sup>Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark

**Objective:** To link the items in the Patient-Reported Outcome Measures (PROMs): Roland Morris Disability Questionnaire, Short Form 36 (SF-36) and pain scores, to the Brief International Classification of Functioning, Disability and Health (ICF) Core Set for low back pain, and to examine the extent to which a clinician's assessment of patients' problems according to the Brief ICF Core Set correlates with the scores of matching items from the PROMs.

**Methods:** The PROMs were linked to the Brief ICF Core Set for low back pain. Secondly, a cross-sectional study was conducted including 70 patients with low back pain. The patients completed the PROMs, and the Brief ICF Core Set for low back pain was assessed by a clinician using qualifiers.

**Results:** The items in the PROMs were successfully linked to the ICF. Twelve of the 38 unique ICF categories derived from the PROMs were covered by the Brief ICF Core Set (34%). A weak correlation was found between the patients' responses and the clinician's assessment.

**Conclusion:** The selected PROMs do not cover the prototypical spectrum of problems encountered in patients with low back pain as defined by the Brief ICF Core Set. The clinical assessment of patients' problems according to the Brief ICF Core Set does not correlate with the scores of matching items from the PROMs.

**Key words:** low back pain; assessment; patient outcome; ICF.

J Rehabil Med 2016; 48: 618–624

*Correspondence address:* Charlotte Ibsen, Department of Public Health, Section of Clinical, Social Medicine and Rehabilitation, Aarhus University and DEFACTUM, Central Denmark Region, Denmark, P.P. Oerums Gade 11, 1B, DK-8000 Aarhus C, Denmark. E-mail: chaibs@rm.dk

Accepted Mar 18, 2016; Epub ahead of print May 13, 2016

## INTRODUCTION

Low back pain (LBP) is a major global public health problem and the leading contributor to years lived with disability (1). Although

patients have different diagnoses, their problems manifest in similar ways to pain and functional limitations (2). LBP is a complex biopsychosocial phenomenon; complete differential diagnosis and assessment is difficult and needs to be comprehensive (3). Substantial work has already been performed in the field of LBP regarding the design of outcome metrics, and clinicians and researchers have used well-validated outcome measure tools, such as the pain scales (visual analogue scale), Roland Morris Disability Questionnaire (RMDQ), the Oswestry Disability Index (ODI), and the Short Form 36 (SF-36) to identify, quantify and assess LBP-related symptoms and functioning (4–6). The use of Patient-Reported Outcome Measures (PROMs) in clinical research is growing. To select appropriate PROMs it is important to determine what should be measured (7). Consensus-based efforts have been made to define sets of outcome measures for research purposes (3, 5). However, a universal set of outcomes to be measured and compared as part of standard clinical practice in LBP has not yet been developed (8). One of the challenges associated with the use of existing outcome measures is the lack of standardization, making comparison of studies difficult (5, 6).

The World Health Organization (WHO) has developed the International Classification of Functioning, Disability and Health (ICF) to provide a common language and framework for addressing the impact of a given health condition on human functioning (9, 10). To enhance the applicability of the ICF in clinical practice, a Comprehensive and a Brief ICF Core Set for patients with LBP has been developed (10, 11).

The purpose of this study was to evaluate the extent to which certain PROMs used in clinical practice (RMDQ, SF-36 and pain scores) cover and match the categories in the Brief ICF Core Set for patients with LBP. The specific aims were: (i) to link the items in the PROMs to the detailed classification in the ICF and examine the extent to which the identified categories in the PROMs covered the Brief ICF Core Set for LBP, and (ii) to examine to what extent a clinician's view on patients' problems according to the Brief ICF Core Set correlates with the scores of matching items from the PROMs.

METHODS

*Design and setting*

Two different methods were applied. First, a methodological study was conducted to link the PROMs to the ICF. Secondly, a cross-sectional study was conducted, in which patients with LBP completed PROMs and clinicians carried out assessments using the Brief ICF Core Set.

*Patients*

Patients with long-lasting LBP (> 3 months) referred for specialized evaluation at the Department of Rheumatology at Aarhus University Hospital, Denmark were consecutively enrolled during a period of 6 months in 2006. Inclusion criteria were: age range 18–55 years. Exclusion criteria were: radiation of pain to the legs, or inability to speak and understand Danish. The study was notified to the committee on biomedical research ethics and all patients provided written informed consent.

*Assessments*

PROMs were completed by the patients before the specialist evaluation. The PROMs used were already being implemented in the Department of Rheumatology. Furthermore, the ICF categories in the Brief ICF Core Set served as standard for clinicians’ assessment, and were therefore not useful to patients.

*Activity limitation.* The LBP-specific questionnaire RMDQ was used to evaluate activity limitations (12). The version of RMDQ used here included 23 questions on activity limitations due to LBP with dichotomous answers “yes” (1 point) or “no” (0 points) (13). Thus, the scale has a score ranging from 0 to 23, with higher scores reflecting greater activity limitation.

*Health-related quality of life.* The generic questionnaire SF-36 was used to evaluate health-related quality of life (14). SF-36 includes 36 items that measure health in 8 subscales: Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. The results on the 8 scales are summarized on a Physical Composite Score (PCS) and a Mental Composite Score (MCS). The scale has a total score from 0 to 100. Low scores indicate that the person has a severely limited scope of activity, while high scores indicate that the person performs all types of tasks. We used the Danish version of the SF-36 (15).

*Pain.* Two vertical visual analogue scales (VAS) were used to evaluate current pain intensity in the lower back and legs (16). Each of the scales consisted of a 100-mm line, anchored from 0 (no pain) to 100 (worst imaginable pain). The patients answered 2 questions: “On a scale from 0 to 100, how much pain are you in right now in your lower back and legs, respectively?”.

In connection with the specialist evaluation each patient was assessed using the Brief ICF Core Set for LBP by a clinician trained in the assessment procedure. The clinician was blinded to the PROM results (11). The Brief ICF Core Set for LBP includes as few categories as possible, but still enough to adequately describe the most important spectrum of problems related to functioning (10, 11). It is less time-consuming compared with the Comprehensive ICF Core Set, and more applicable to specialist evaluation in clinical practice (10, 11, 17, 18). The Brief ICF Core Set includes 35 2<sup>nd</sup> level ICF categories: 10 for “body functions”, 3 for “body structure”, 12 for “activities and participation” and 10 for “environmental factors” (19). The ICF categories identified during the linking process were compared with categories in the Brief ICF Core Set for LBP. The identified 3<sup>rd</sup> and 4<sup>th</sup> level categories (e.g. d4500 “walking short distances”) were reassigned to the cor-

responding 2<sup>nd</sup> level category (d450 Walking), as the Brief ICF Core Set for LBP includes only 2<sup>nd</sup> level categories. The generic qualifier scale was used to describe the patient’s problem in each ICF category. The qualifier scale of the components “body functions”, “body structure” and “activities and participation” has 5 response levels, ranging from 0 to 4: “no”, “mild”, “moderate”, “severe” and “complete problem”. In addition, the qualifiers “8” (not specified) and “9” (not applicable) could be used. In order to ensure consistency and reproducibility of the Brief ICF Core Set assessment, only patients assessed by the same clinician were included. In this study we assessed the components “body functions”, “body structure” and “activities and participation”. The environmental factors were not assessed, as they were considered irrelevant in this context.

*Linking procedures*

ICF linking rules make it possible to link concepts contained in outcome measures to ICF in a standardized way (20, 21). Each meaningful concept contained in an item must be identified and linked to the most precise ICF category (2<sup>nd</sup>, 3<sup>rd</sup> or 4<sup>th</sup> level). If an item contains more than one concept, each concept has to be linked separately; this is because a given question can correspond to more than 1 ICF category. If meaningful concepts are not sufficient to link to an ICF category, the concept is assigned “not definable”. Special cases of this are concepts referring to general health, physical health or mental health (20). Our linking process included 61 items from the PROMs: 23 from the RMDQ, 36 from the SF-36 and 2 from pain scales. Meaningful concepts from the items in the PROMs were extracted and linked to the ICF by 2 independent researchers using the linking rules (20). In case of disagreement, consensus was reached through discussion; in the case of persistent disagreement, a third independent researcher was consulted on the most suitable category.

*Statistical analysis*

Statistical analysis was performed using STATA 13 and non-parametric tests were used due to ordinal data and the non-normality of the data. SF-36 scores were computed using rough scores; the total sum score of RMDQ (0–23 point) was calculated, and scores from the 2 pain scales were analysed separately. Missing responses from the PROMs were excluded from the analysis. There were no missing responses from the Brief ICF Core Set, but qualifiers “8” and “9” were analysed as missing. The correlation between the clinician’s assessment and the patient’s responses were analysed by Spearman’s correlations (*r*). Correlation coefficients at  $\geq 0.7$  were considered good (22). Results are presented as medians (IQR).

RESULTS

The linking process identified 104 meaningful concepts within the 3 PROMs. Out of these 104 concepts, 75 (72%) could be linked to the ICF, while 29 (28%) could not (Table I).

Table I. Number of meaningful concepts identified in the Patient-Reported Outcome Measures (PROMs) and numbers linked to the International Classification of Functioning, Disability and Health (ICF)

	RMDQ n (%)	SF-36 n (%)	Pain n (%)	Total n (%)
Number of items	23	36	2	61
Number of meaningful concepts	34	68	2	104
Meaningful concepts linked to an ICF category	26 (76)	47 (69)	2 (100)	75 (72)
Meaningful concepts not linked to an ICF category	8 (24)	21 (31)	0	29 (28)
Not definable	7	10	0	17
Not definable – general health	1	8	0	9
Not definable – physical health	0	3	0	3
Not definable – mental health	0	0	0	0

RMDQ: Roland Morris Disability Questionnaire.

Table II. *International Classification of Functioning, Disability and Health (ICF) Categories represented in the Roland Morris Disability Questionnaire (RMDQ), Short Form 36 (SF-36), pain scores and the Brief ICF Core Set for low back pain (LBP)*

ICF category (level)	ICF category title	RMDQ	SF-36	Pain	Brief ICF Core Set
2 <sup>nd</sup> 3 <sup>rd</sup> 4 <sup>th</sup>					
<i>Body functions</i>					
b130	Energy and drive functions		X		X <sup>a</sup>
b134	Sleep functions				X <sup>a</sup>
b1343	Quality of sleep	X			
b152	Emotional functions	X	X		X <sup>a</sup>
b280	Sensation of pain	X	X		X <sup>a</sup>
b2801	Pain in body part	X			
b28013	Pain in back	X		X	
b28015	Pain in lower limb	X		X	
b455	Exercise tolerance functions				X
b710	Mobility of joint functions				X
b715	Stability of joint functions				X
b730	Muscle power functions				X
b735	Muscle tone functions				X
b740	Muscle endurance functions				X
<i>Body structures</i>					
s120	Spinal cord and related structures				X
s760	Structure of trunk				X
s770	Additional musculoskeletal structures related to movement				X
<i>Activities and participation</i>					
d230 <sup>b</sup>	Carrying out daily routine		X		
d240	Handling stress and other psychological demands				X
d410	Changing basic body position	X			X <sup>a</sup>
d4101	Squatting	X	X		
d4103	Sitting	X			
d4105	Bending	X	X		
d4106	Shifting the body's centre of gravity	X			
d415	Maintaining a body position				X
d4154	Maintaining a standing position	X			
d430	Lifting and carrying objects		X		X <sup>a</sup>
d450	Walking	X			X <sup>a</sup>
d4500	Walking short distances	X	X		
d4501	Walking long distances		X		
d451 <sup>b</sup>	Climbing	X	X		
d4552	Running		X		
d4750 <sup>b</sup>	Driving human-powered transportation		X		
d510 <sup>b</sup>	Washing oneself		X		
d530	Toileting				X
d540	Dressing	X	X		X <sup>a</sup>
d5402	Putting on footwear	X			
d5700 <sup>b</sup>	Ensuring one's physical comfort	X			
d620 <sup>b</sup>	Acquisition of goods and services		X		
d640	Doing housework	X	X		X <sup>a</sup>
d650 <sup>b</sup>	Caring for household objects	X	X		
d7202 <sup>b</sup>	Regulating behaviours within interactions	X			
d760	Family relationships		X		X <sup>a</sup>
d7702 <sup>b</sup>	Sexual relationships	X			
d845	Acquiring, keeping and terminating a job				X
d850	Remunerative employment		X		X <sup>a</sup>
d859	Work and employment, other specified and unspecified				X
d920 <sup>b</sup>	Recreation and leisure		X		
d9201	Sports		X		
d9205	Socializing		X		
<i>Environmental factors</i>					
e110	Products or substances for personal consumption				X
e1150 <sup>b</sup>	General products and technology for personal use in daily living	X			
e135	Products and technology for employment				X
e150 <sup>b</sup>	Design, construction and building products and technology of buildings for public use	X			
e155	Design, construction and building products and technology of buildings for private use	X			X <sup>a</sup>
e310	Immediate family				X
e355	Health professionals				X

Table II. *Contd.*

ICF category (level)	ICF category title	RMDQ	SF-36	Pain	Brief ICF Core Set
e410	Individual attitudes of immediate family members				X
e450	Individual attitudes of health professionals				X
e550	Legal services, systems and policies				X
e570	Social security services, systems and policies				X
e580	Health services, systems and policies				X

PROMs: Patient-Reported Outcome Measures. X<sup>a</sup>: ICF category derived from the PROMs and included in The Brief ICF Core Set; <sup>b</sup>ICF categories derived from the PROMs and not represented in The Brief ICF Core Set.

The linking process showed that several items in both the RMDQ and the SF-36 addressed more than 1 ICF category and that 1 ICF category was represented more than once in the PROMs. Due to overlaps in the PROMs, the 75 meaningful concepts were linked to 49 ICF categories. There were 11 overlapping ICF categories, which meant that a total of 38 unique ICF categories were derived from the PROMs (Table II). Of the 38 ICF categories, 17 were 2<sup>nd</sup> level, 19 were 3<sup>rd</sup> level and 2 were 4<sup>th</sup> level. The Brief ICF Core Set includes 35 categories at 2<sup>nd</sup> level, yet only 12 categories were covered by items in the PROMs (34%) (X<sup>a</sup> in Table II), leaving 23 ICF categories not covered by the PROMs. On the other hand, 12 ICF categories derived from the PROMs were not represented in the Brief ICF Core Set. Details can be requested from the corresponding author.

In total, 96 patients were referred for specialist evaluation at the Department of Rheumatology and 70 were included in the cross-sectional study (Fig. 1). Baseline characteristics are presented in Table III.

The correlation analysis comprised the 12 ICF categories from “body functions” and “activities and participation” (Table IV). Results from the RMDQ showed statistically significant correlations ( $r_s$ ) between the analysed item and the corresponding ICF category for 7 out of 17 items. For the SF-36, statistically significant correlations were found in 24 of 33 items. A correlation was also found between the item concerning current pain intensity

(VAS) in the lower back and the corresponding ICF category b280 “sensation of pain”, but not for the item concerning current pain (VAS) in the legs. None of the identified correlation coefficients reached the threshold value of 0.7 (range -0.1 to 0.6).

### DISCUSSION

This study provides important insight into whether these PROMs cover and match the categories in the Brief ICF Core Set for patients with LBP, and to what extent a clinician’s view on their patients’ problems correlates with those patients’ scoring of matching items from the PROMs.

The linking result showed that the items in the PROMs were successfully linked to the detailed classification in the ICF, and that 12 of the ICF categories derived from the PROMs are covered by the Brief ICF Core Set for LBP. The cross-sectional study showed a low correlation between the clinician’s assessment using the Brief ICF Core Set and the patient’s scores of the matching items from the PROMs.

Our linking result is in accordance with results from previous studies (10, 21, 23, 24). One other study used RMDQ, SF-36 and ICF Core Set for patients with LBP, but different aims and statistical methods made it impossible to compare the results (25).

A strength of this study is that the selected PROMs measure the outcome domains recommended in a core outcome set for

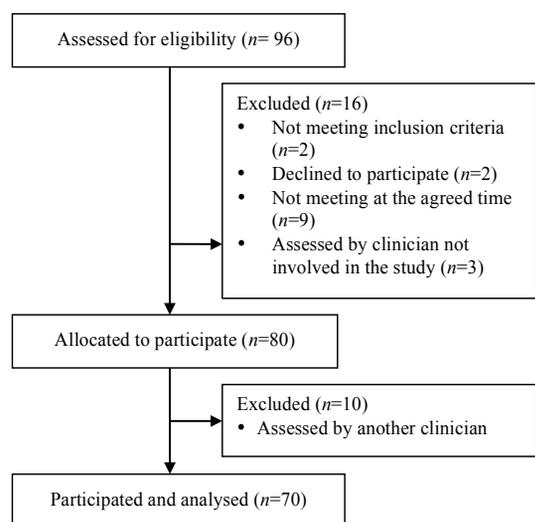


Fig. 1. Study participation.

Table III. *Baseline characteristics of participants (n = 70)*

Gender, female, n (%)	42 (60)
Age, years, mean (SD)	36.6 (8.9)
Scores of patients on PROMs (range)	
VAS-pain in lower back <sup>a</sup> (0–100), median (IQR)	43 (27–59)
VAS-pain in leg <sup>b</sup> (0–100)	15 (1–50)
RMDQ total sum score (0–23), median (IQR)	14 (8–17)
SF-36 (0–100), median (IQR)	
Physical functioning	65 (50–80)
Role-physical <sup>a</sup>	25 (0–50)
Bodily pain	32 (22–41)
General health	55 (35–67)
Vitality	45 (30–60)
Social functioning	75 (50–100)
Role-emotional <sup>c</sup>	100 (66.7–100)
Mental health	72 (56–80)
Physical health summary component <sup>c</sup>	32.5 (27.5–41)
Mental health summary component <sup>c</sup>	52.3 (42.8–59.4)

<sup>a</sup>n = 69, <sup>b</sup>n = 67, <sup>c</sup>n = 66.

PROMs: Patient-Reported Outcome Measures; RMDQ: Roland Morris Disability Questionnaire; VAS: visual analogue scale.

Table IV. Correlations between International Classification of Functioning, Disability and Health (ICF) categories in the Brief ICF Core Set for low back pain (LBP) and items in the Patient-Reported Outcome Measures (PROMs) (n = 70)

Brief ICF Core Set for LBP		RMDQ			SF-36			Pain		
ICF code (2 <sup>nd</sup> level)	ICF category title	Item no	$r_s$	$p$ -value	Item no	$r_s$	$p$ -value	Item no	$r_s$	$p$ -value
b130	Energy and drive functions				23	0.4 <sup>g</sup>	0.0010			
					27	0.4 <sup>g</sup>	0.0004			
					29	0.3 <sup>h</sup>	0.0385			
					31	0.3 <sup>h</sup>	0.0040			
b134	Sleep functions	15	0.4	0.0010						
b152	Emotional functions	17	0.1 <sup>e</sup>	0.3246	17	0.1 <sup>a</sup>	0.3472			
					19	0.0 <sup>a</sup>	0.8208			
		23	-0.1 <sup>e</sup>	0.6481	20	0.2 <sup>e</sup>	0.0664			
					23	0.4 <sup>d</sup>	0.0043			
					24	0.1 <sup>e</sup>	0.4474			
					25	0.0 <sup>d</sup>	0.7711			
					26	0.3 <sup>d</sup>	0.0272			
					28	0.3 <sup>e</sup>	0.0386			
					30	0.3 <sup>e</sup>	0.0078			
					32	0.2 <sup>e</sup>	0.1475			
b280	Sensation of pain	11	0.3	0.0088	21	0.4 <sup>i</sup>	0.0006	Low back	0.4 <sup>9</sup>	0.0008
d410	Changing basic body position	2	0.2 <sup>i</sup>	0.1089	8	0.4	0.0014	Leg	0.1 <sup>7</sup>	0.4583
		6	0.1 <sup>i</sup>	0.4083						
		9	0.2 <sup>i</sup>	0.0625						
		10	0.1	0.2644						
		12	0.3	0.0266						
d415	Maintaining a body position	8	0.0	0.9019						
d430	Lifting and carrying objects				3	0.4 <sup>f</sup>	0.0002			
					4	0.4 <sup>e</sup>	0.0011			
					5	0.4 <sup>f</sup>	0.0004			
d450	Walking	3	0.4	0.0010	9	0.4	0.0001			
		14	0.3	0.0297	10	0.3	0.0059			
					11	0.3	0.0034			
d540	Dressing	7	0.2	0.0613	12	0.2	0.0793			
		13	0.3 <sup>i</sup>	0.0108						
d640	Doing housework	4	0.2 <sup>h</sup>	0.0907	4	0.4 <sup>d</sup>	0.0003			
		16	0.3 <sup>h</sup>	0.009	22	0.5 <sup>h</sup>	0.0000			
		22	0.2 <sup>h</sup>	0.0895						
d760	Family relationships				20	0.4 <sup>b</sup>	0.0004			
d850	Remunerative employment				13	0.4 <sup>c</sup>	0.0004			
					15	0.4 <sup>d</sup>	0.0004			
					16	0.5 <sup>c</sup>	0.0001			
					17	0.2 <sup>d</sup>	0.0961			
					19	0.2 <sup>d</sup>	0.2152			
			22	0.5 <sup>h</sup>	0.0000					

PROMs: Patient-Reported Outcome Measures, RMDQ: Roland Morris Disability Questionnaire; SF-36: Short Form 36. Correlations calculated by Spearman's correlations ( $r$ );  $r$  is statistically significant  $p \leq 0.05$ , <sup>a</sup> $n=61$ , <sup>b</sup> $n=62$ , <sup>c</sup> $n=63$ , <sup>d</sup> $n=64$ , <sup>e</sup> $n=65$ , <sup>f</sup> $n=66$ , <sup>g</sup> $n=67$ , <sup>h</sup> $n=68$ , <sup>i</sup> $n=69$ .

patients with LBP; they are well-validated and have been recommended since 1998 (3, 5). RMDQ is the most commonly used questionnaire for measuring activity limitation. The SF-36 is a frequently used PROM in clinical trials, and VAS is recommended as a valid and reliable instrument to measure pain (8, 26–28). To ensure consistent ICF assessment we trained the clinicians in the use and understanding of the Brief ICF Core Set and the qualifiers. In the analysis, we only included the assessment from 1 clinician in order to ensure consistency in the assessment. The Brief ICF Core Set was in itself developed over the course of a comprehensive process; it has been tested in different LBP populations and is suitable to provide a general overview of functional limitations (11, 18, 23, 29–31). The Brief ICF Core Set was also considered most appropriate for

use in a busy clinical setting with limited time for examination, as it would impose the smallest possible time burden on both patients and clinicians. This concurs with other studies, which found the Comprehensive Core Set to be too time-consuming for clinical use (17, 18, 32). If the Comprehensive Core Set were used as a reference, the PROMs would have covered 21 of 78 ICF categories (27%), resulting in a smaller coverage than the Brief ICF Core Set (34%). Thus, we still recommend the Brief ICF Core Set to be the most appropriate in clinical practice.

This study has some limitations. First of all the assumption that the problems identified by the PROMs and the problems identified by a clinician would be similar. Even though the items in the PROMs and the Brief ICF Core Set pertain to the same ICF category, they are formulated and scored very differently;

thus, differences related to interpretation are likely. Moreover, patients' and clinicians' perceptions of problems may differ despite use of the exact same outcome measures. The Brief ICF Core Set has been criticised for not providing sufficient clinical information, as the information is not helpful in individually-tailored treatment planning (30). To include more of the common patient-reported limitations, studies suggest adding 3 categories to the Brief ICF Core Set (29, 30). The categories: "recreation and leisure" (d920) and "caring for household objects" (d650) are covered in SF-36, while the category "pulling/pushing" (d445) is not covered in the PROMs. Adding these categories would probably not have affected our results. Another limitation arises when Brief ICF Core Set categories are present at the 2nd level, while several items in the PROMs are linked to 3rd and 4th level categories (20 of 38 categories) (Table II). We reassigned 3rd and 4th level ICF categories to 2nd level categories to match the categories in the Brief ICF Core Set. This reassignment does not allow the same kind of specific and nuanced description of functioning that has also been reported in other studies (30). We used the Brief ICF Core Set as a reference; another option would have been to use the PROMs as reference. This would add another 12 categories (in Table II) to the clinical assessment. As the SF-36 is a generic outcome measure, it is difficult to determine whether all of the 7 additional categories from the SF-36 would be clinically relevant in LBP. Using RMDQ and pain scores as references would have added another 4 categories. The choice of reference does not change the correlation coefficients; it would just add further categories. We believe that the additional categories would show the same pattern and thus not alter the general results. Finally, using generic qualifiers as measurement instruments has been criticised and a revision is recommended (33, 34). This is supported by our findings as the clinician only used "complete problems" ("4") or "severe problems" ("3") a few times. In our analysis, the qualifier "not specified" ("8") was applied as missing, which may have affected our correlation result. Additional analyses, with "8" interpreted as "worst case" and replaced with "4" (complete problem), showed only minor, non-significant changes in the results. A dichotomized test could have been applied to test whether the PROMs could cover the clinician's identification of problems. Additional analysis on dichotomized data from the PROMs and clinician-identified problems showed no correlation.

RMDQ has recently been criticized due to some misfitting items and poor targeting of items with high and low activity limitations (35). The study concludes that it is time to consider a reconstruction of RMDQ using an approach based on item response theory; this should include considerations of new items and response options. RMDQ is often suggested as a measurement instrument for LBP, but in the newly updated guidelines no instrument is recommended (3, 5). Instead, it is advised that the next step is to reach consensus on which instruments to use (3). The SF-36 is still very widely used in spite of many hundred PROMs being used in clinical trials (27). Our study suggests that the SF-36 does not provide sufficient information on the individual patient's health-related quality of life. During our linking process, we found that several items in the SF-36

contained more than 1 ICF category. Item number 20 in SF-36: "Have your physical health or emotional problems interfered with your social activities with family, friends, neighbours, or groups?" – is an example: it is linked to both "emotional functions" (b152) and "family relations" (d760) (Table IV). These 2 ICF categories describe 2 very diverse constructs, and in the ICF terminology they relate to 2 different components; "body functions" and "activities and participation". This was supported by our analysis, showing significant correlation in "emotional functions" (b152) and no significance in "family relations" (d760). This example may support the conclusion that the SF-36 has been developed to target health conditions at group level, not at individual patient level, which may explain why the SF-36 has little value for clinicians when examining individual patients. Hence, we suggest that it is time to develop new instruments to measure individual health-related quality of life.

Measuring outcomes in patients with LBP is challenging, as numerous diseases or conditions can result in LBP (8). In addition, LBP can rarely be measured by objective endpoints, and outcomes are therefore often measured with PROMs, which are subjective and require thorough psychometric testing (8). The International Consortium for Health Outcomes Measurement (ICHOM) recently presented a standard set of outcome metrics to assess management of LBP in clinical practice (8). At the same time, ICHOM proclaimed that revisions to this outcome set are needed, as new software enables the use of computerized adaptive testing (CAT) in PROM collection (8). An example of the use of CAT is the Patient-Reported Outcomes Measurement Information System (PROMIS®). PROMIS instruments have been developed using psychometric methods, including factor analysis and item response theory, to provide precise and efficient measurement of the underlying concept (36, 37). PROMIS instruments have been linked to the ICF, and the ICF can serve as a typology for classifying the content of PROMIS (38, 39). This adds valuable knowledge to our findings and it would be interesting to examine whether PROMIS instruments could be the answer to some of the problems concerning the PROMs used in this study. However, when clinicians and researchers are deciding on the right instrument for measuring a specific endpoint, it is essential to determine what to measure and then to know which domains will be covered. Our study contributes to this by revealing which domains the selected PROMs cover and which they do not.

We found that these PROMs (RMDQ, SF-36 and pain scores) can be successfully linked to the detailed classification in the ICF, and that ICF categories derived from the PROMS cover 34% of the categories in the Brief ICF Core Set for LBP. We found weak correlations between the problems identified by the PROMs and problems identified by a clinician. If the Brief ICF Core Set is used as a reference, PROMs are imprecise, as several items target more than 1 ICF category. Thus, we conclude that the PROMs do not cover the prototypical spectrum of problems encountered in patients with LBP as defined by the Brief ICF Core Set. Our study showed that a clinician's view on patients' problems according to the Brief ICF Core Set does not correlate with the scores of matching items from the PROMs.

## REFERENCES

- Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014; 73: 968–974.
- Andersson GB. Epidemiology of low back pain. *Acta Orthop Scand Suppl* 1998; 281: 28–31.
- Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J* 2015; 24: 1127–1142.
- Burton AK, Waddell G. Clinical guidelines in the management of low back pain. *Baillieres Clin Rheumatol* 1998; 12: 17–35.
- Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine (Phila Pa 1976)* 1998; 23: 2003–2013.
- 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.
- Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res* 2013; 22: 2265–2278.
- Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop* 2015; 86: 523–533.
- World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: WHO; 2001.
- Stier-Jarmer M, Cieza A, Borchers M, Stucki G, World Health Organization. How to apply the ICF and ICF core sets for low back pain. *Clin J Pain* 2009; 25: 29–38.
- Cieza A, Stucki G, Weigl M, Disler P, Jackel W, van der Linden S, et al. ICF Core Sets for low back pain. *J Rehabil Med* 2004; Suppl 44: 69–74.
- Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine (Phila Pa 1976)* 1995; 20: 1899–908; discussion 1909.
- Albert HB, Jensen AM, Dahl D, Rasmussen MN. Criteria validation of the Roland Morris questionnaire. A Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica. *Ugeskr Laeger* 2003; 165: 1875–1880.
- Ware JE, Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–483.
- Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: translation and preliminary validity studies. *J Clin Epidemiol* 1998; 51: 991–999.
- Maxwell C. Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment. *Br J Clin Pharmacol* 1978; 6: 15–24.
- Roe C, Sveen U, Kristoffersen OJ, Fossen B, Hammergren N, Iversen VT, et al. Testing of ICF core set for low back pain. *Tidsskr Nor Laegeforen* 2008; 128: 2706–2708.
- Roe C, Sveen U, Cieza A, Geyh S, Bautz-Holter E. Validation of the Brief ICF core set for low back pain from the Norwegian perspective. *Eur J Phys Rehabil Med* 2009; 45: 403–414.
- ICF Research Branch. ICF Core Set for Low Back Pain. 2013; [accessed 2015 Apr 21]. Available at: [http://www.icf-research-branch.org/images/stories/ICF%20Core%20Sets%20Download/Comprehensive\\_and\\_Brief\\_ICF\\_Core\\_Sets\\_Low\\_Back\\_Pain.pdf](http://www.icf-research-branch.org/images/stories/ICF%20Core%20Sets%20Download/Comprehensive_and_Brief_ICF_Core_Sets_Low_Back_Pain.pdf).
- Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. *J Rehabil Med* 2005; 37: 212–218.
- Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health. *J Rehabil Med* 2002; 34: 205–210.
- Dancey CP. Statistics without maths for psychology: using SPSS for Windows. 3rd edn. Harlow: Prentice Hall; 2004.
- Wang P, Zhang J, Liao W, Zhao L, Guo Y, Qiu Z, et al. Content comparison of questionnaires and scales used in low back pain based on the international classification of functioning, disability and health: a systematic review. *Disabil Rehabil* 2012; 34: 1167–1177.
- Sigl T, Cieza A, Brockow T, Chatterji S, Kostanjsek N, Stucki G. Content comparison of low back pain-specific measures based on the International Classification of Functioning, Disability and Health (ICF). *Clin J Pain* 2006; 22: 147–153.
- Öztuna D, Yanik B, Kutlay Ş, KurtalŞ Aytür Y, Elhan AH, Tennant A, et al. Psychometric properties of the ICF Core Set for Low Back Pain and its clinical use. *Turkish J Rheumatol* 2011; 26: 44.
- Froud R, Eldridge S, Kovacs F, Breen A, Bolton J, Dunn K, et al. Reporting outcomes of back pain trials: a modified Delphi study. *Eur J Pain* 2011; 15: 1068–1074.
- Scoggins JF, Patrick DL. The use of patient-reported outcomes instruments in registered clinical trials: evidence from ClinicalTrials.gov. *Contemp Clin Trials* 2009; 30: 289–292.
- Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. *J Pain* 2006; 7: 192–199.
- Mullis R, Barber J, Lewis M, Hay E. ICF core sets for low back pain: do they include what matters to patients? *J Rehabil Med* 2007; 39: 353–357.
- Lygren H, Strand LI, Anderson B, Magnussen LH. Do ICF core sets for low back pain include patients' self-reported activity limitations because of back problems? *Physiother Res Int* 2014; 19: 99–107.
- Paul B, Leitner C, Vacariu G, Wick F, Zehetmayer S, Matzner M, et al. Low-back pain assessment based on the Brief ICF Core Sets: diagnostic relevance of motor performance and psychological tests. *Am J Phys Med Rehabil* 2008; 87: 452–460.
- Bautz-Holter E, Sveen U, Cieza A, Geyh S, Roe C. Does the International Classification of Functioning, Disability and Health (ICF) core set for low back pain cover the patients' problems? A cross-sectional content-validity study with a Norwegian population. *Eur J Phys Rehabil Med* 2008; 44: 387–397.
- Uhlig T, Lillemo S, Moe RH, Stamm T, Cieza A, Boonen A, et al. Reliability of the ICF Core Set for rheumatoid arthritis. *Ann Rheum Dis* 2007; 66: 1078–1084.
- Roe C, Sveen U, Geyh S, Cieza A, Bautz-Holter E. Construct dimensionality and properties of the categories in the ICF Core Set for low back pain. *J Rehabil Med* 2009; 41: 429–437.
- Kent P, Grotle M, Dunn KM, Albert HB, Lauridsen HH. Rasch analysis of the 23-item version of the Roland Morris Disability Questionnaire. *J Rehabil Med* 2015; 47: 356–364.
- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007; 45: S3–S11.
- Riley WT, Rothrock N, Bruce B, Christodolou C, Cook K, Hahn EA, et al. Patient-reported outcomes measurement information system (PROMIS) domain names and definitions revisions: further evaluation of content validity in IRT-derived item banks. *Qual Life Res* 2010; 19: 1311–1321.
- Tucker CA, Cieza A, Riley AW, Stucki G, Lai JS, Bedirhan Ustun T, et al. Concept analysis of the patient reported outcomes measurement information system (PROMIS(R)) and the international classification of functioning, disability and health (ICF). *Qual Life Res* 2014; 23: 1677–1686.
- Tucker CA, Escorpizo R, Cieza A, Lai JS, Stucki G, Ustun TB, et al. Mapping the content of the Patient-Reported Outcomes Measurement Information System (PROMIS(R)) using the International Classification of Functioning, Health and Disability. *Qual Life Res* 2014; 23: 2431–2438.