ORIGINAL REPORT

DEVELOPMENT OF A PATIENT-REPORTED OUTCOME: THE NECK OUTCOME SCORE (NOOS) – CONTENT AND CONSTRUCT VALIDITY

Tina Juul, MHSc[1,2], Karen Søgaard, PhD[1], Ewa M. Roos, PhD[1] and Aileen M. Davis, PhD[3]

From the [1]Institute of Sports Science and Clinical Biomechanics, The University of Southern Denmark, Odense, [2]CopenRehab, Section of Social Medicine, Department of Public Health, University of Copenhagen, Denmark and [3]Division of Health Care and Outcomes Research, Toronto Western Research Institute, University Health Network and the University of Toronto, Toronto, Canada

Objective: **To develop a patient-reported outcome evaluating the impact of neck pain. The results of item generation and reduction and subscale structure in support of the content and construct validity of the measure are reported.**
Methods: **Items were generated from the literature and through focus groups including patients with neck pain and healthcare professionals, respectively. Item reduction was based on focus groups and field-tested questionnaire data. Construct validity was assessed using exploratory factor analysis.**
Results: **Focus groups containing 24 patients (mean age 57.2 (standard deviation (SD) 15.9) years, range 24–85 years); 19 women) and 12 healthcare professionals were conducted before data saturation was achieved. A total of 196 patients with neck pain (mean age 47.8 (SD 13.7) years), range 18–89 years; 146 women) completed the preliminary questionnaire. Overall 35 items were removed from the original 69. A multidimensional questionnaire, divided into five subscales, was developed from the remaining 34 items: mobility; symptoms; sleep disturbance; everyday activity and pain; and participation in everyday life. Exploratory factor analysis supported a 5-subscale structure.**
Conclusion: **The Neck OutcOme Score has excellent content validity and preliminary results support a 5-subscale structure. Additional work is needed to assess the reliability, further construct validity and responsiveness.**

Key words: patient-reported outcome; questionnaire; content validity; construct validity; neck pain.

INTRODUCTION

Neck pain (NP) is the second most common musculoskeletal disorder (1). At any given time 12–14% of the adult population reports having NP (2), which has been related to neuromuscular deficits (3), deficits in eye movement control (4) and impaired performance of muscles in the cervical spine (5). As a result,

NP often causes reduced work ability and significantly increases the risk of long-term sickness absence (6). Given the burden of NP, there is an urgent need for efficacious interventions. However, evaluation of effectiveness and efficacy require salient and robust outcome measures.

There is a lack of patient-reported outcomes (PROs) with sound psychometric properties for the evaluation of patients with NP. The Neck Disability Index (NDI) (7), one of the most applied PROs for NP, has been criticized regarding content validity (8), reliability (9, 10) and dimensionality (10). The NDI was originally developed to measure disability in activities of daily living (7). However, the development process of NDI presents shortcomings, such as a broad selection of patients, achievement of data saturation and lack of patient input in determining content validity (11). This is mainly due to the absence of clear standardized procedures (12). While current measures can be modified and their measurement properties evaluated for people with NP, the need for patient input to determine content validity (12) based on a conceptual framework (13) is often lacking or inadequate. We therefore elected to develop a PRO that evaluates the impact of NP related to impairment, activity and participation, the Neck OutcOme Score (NOOS). Prior to item generation, the concepts being measured need to be clearly defined. The International Classification of Functioning, Disability and Health (ICF) (14) bio-psycho-social framework is commonly used for the development of PROs (15) and the efficacy of a given treatment is often evaluated based on content related to the classification codes (16). The ICF includes 3 domains: body functions and structure, activity, and participation (14). While there is guidance for the content of body functions and structure, the activity and participation domains are combined, leaving the content of the individual domains open to interpretation (17). Nevertheless, the ICF is the most applied framework. The present study reports the results of item generation and reduction and subscale structure in support of the content and construct validity of the NOOS.

METHODS

The study was approved by the Regional Scientific Ethical Committee for Southern Denmark, Denmark and the Danish Data Protection Agency (ID: 30513). All participants provided informed consent, and all rights of the participants were protected.

Questionnaire development

Development of the NOOS was guided by the Consensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (13) recommendations.

Target population identification

In order to represent people with a broad spectrum of NP, eligibility criteria included people over 18 years of age, who were able to communicate in written and spoken Danish, who had NP (with/without cervical radiculopathy) indicated on a body diagram (18), and who sought treatment for their NP. Patients were excluded if they had or were being investigated for cancer; had a neurological disorder (e.g. Multiple Sclerosis); lacked the ability to communicate (e.g. due to a psychotic episode) or had a history of alcohol or drug abuse.

Conceptual model

The development of the NOOS was based on a reflective model (classic test theory) in which the construct is reflected by the items (19). The ICF framework was used as the conceptual model; hence the classification system was used as guidance when assigning a given item a priori to a specific subscale. Body functions, body structure, activity and participation were defined in accordance with the ICF

framework (20). In addition, the NOOS used the template of the Knee injury and Osteoarthritis Outcome Score (KOOS), a multi-dimensional measure with 5 subscales (21). This template (i.e. layout) was chosen, since it has shown good acceptability in patients with ankle, knee, hip and groin-related problems (18, 22, 23). A 5-point ordinal scale was used, with response options ranging from "no problems" (0) to having "extreme problems" (4). A normalized score (0–100, worst to best) is calculated for each subscale, using the following formula: 100–mean (itemX: itemX)/4*100, which provides a patient profile.

Item generation and reduction

Fig. 1 shows an outline of the item generation process and reduction.

Items were generated through a systematic literature review and focus groups with patients and healthcare professionals (HCPs) to ensure content validity. A draft questionnaire was developed by the first author (TJ), grouping items into subscales based on the conceptual model from more than 100 items identified from existing PROs for patients with NP (e.g. 7, 24–33). For example, items regarding range of motion were grouped with similar items concerning body functions and structure (e.g. stiffness and dizziness) to form a subscale. An item such as: *"I have a fair degree of difficulty concentrating"* from the NDI (7) was reformulated to *"Have you had difficulty concentrating?"*. According to the ICF it was classified as body functions and structure
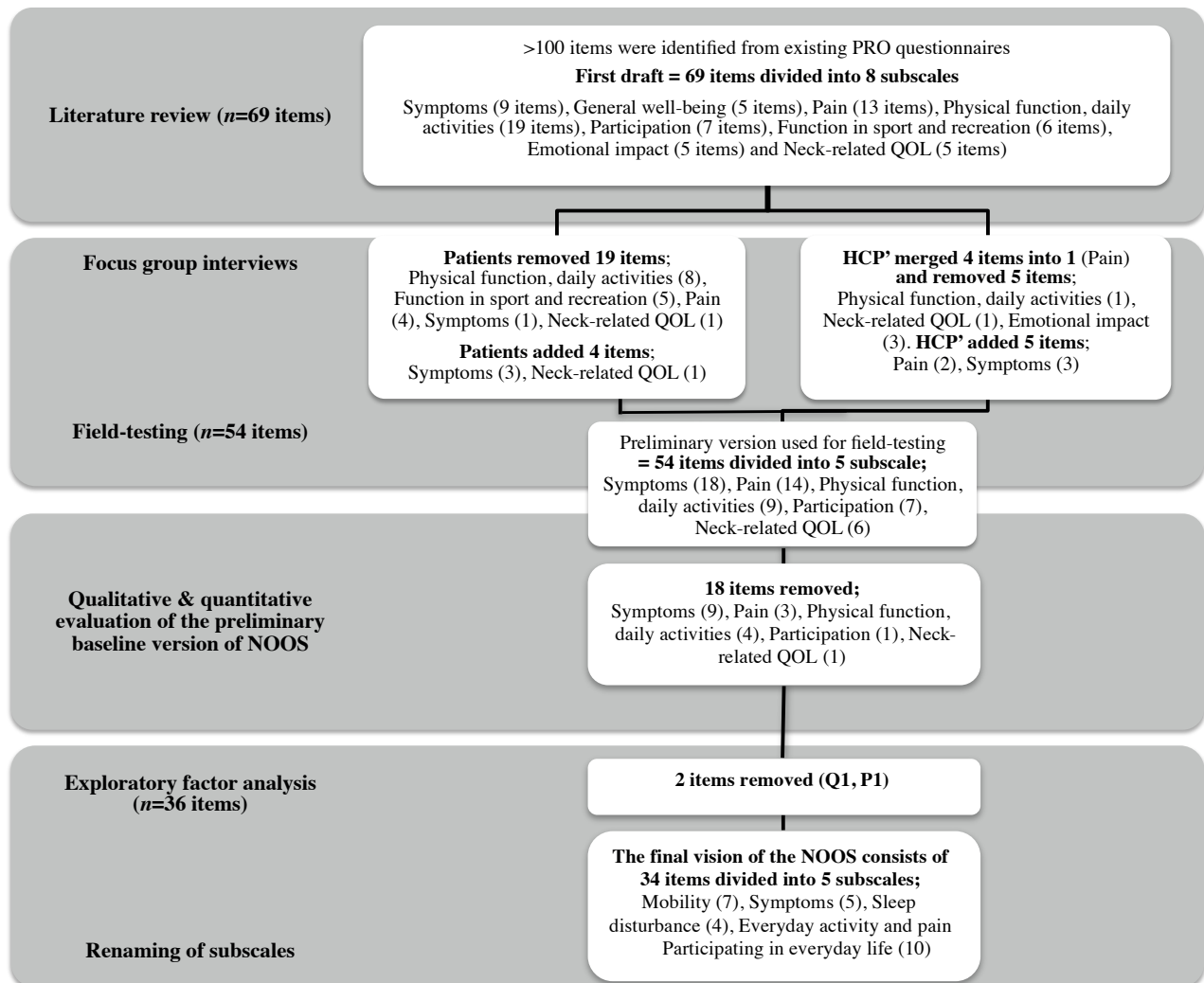

Fig. 1. Item reduction and generation process. PRO: Patient Reported Outcome; QOL: Quality Of Life; NOOS: Neck OutcOme Score; Q1: item 1, neck-related quality of life; P1: item 1, pain subscale; HCP': Health care professionals.

and therefore added to the "symptoms" subscale. Similarly, items were assigned to either a subscale related to activities or participation (e.g. activity items such as shopping or housework, while social and leisure activities formed the participation subscale). Three members of the study group (TJ, EMR, KS) evaluated the draft questionnaire and minor changes were made to item wording. After removing repeated items, the preliminary version of the NOOS, included 69 items with 8 subscales: symptoms (9 items), general well-being (5 items), pain (13 items), physical function/daily activities (19 items), participation (7 items), function in sport and recreation (6 items), emotional impact (5 items) and neck-related quality of life (QOL) (5 items).

Focus groups were then conducted with patients and HCPs to identify additional relevant items and test the comprehensibility, relevance, acceptability and feasibility of the questionnaire (pilot-testing). Patients were recruited from 4 private physiotherapy clinics located in Copenhagen, Aarhus and Odense, Denmark. To ensure that patients were representative of the target population all patients' with NP seeking treatment were included (e.g. patients with whiplash-associated disorder, chronic and acute NP, disc prolapses, derangement, spinal stenosis and osteoarthritis).

The groups of HCPs included physiotherapists specializing in musculoskeletal physiotherapy who treated patients with NP daily and researchers focusing on NP who had experience developing PROs or who had linguistic skills. Each focus group consisted of 2–5 informants with a mix of men and women and was facilitated by the same moderator (TJ). Focus groups were conducted until no new information was elicited (i.e. data saturation).

Focus group participants completed the preliminary version of the questionnaire. They were asked about their experiences with completing the questionnaire and identifying missing content (items) as well as commenting on problem items (e.g. lack of relevance, comprehensibility, language and readability). The patients were asked questions, such as how they interpreted the various terms? How did they choose their answers? What context did they use to answer the item and did they think about their most serious episodes or was it an average? Likewise, they were asked to comment on issues such as whether or not they would be willing to complete a questionnaire, which took approximately 10 min to complete (i.e. acceptability) and whether or not the participants would be able to complete the questionnaire themselves or would need help (feasibility).

Initial item reduction was performed based on focus group data. To ensure it was the patients who perceived a given item as meaningful, all items were included in the focus groups (Fig. 1). An item was removed if none of the patients had experienced the symptoms or had any difficulties with the given activity. Any new items or issues concerning items identified in prior focus groups were discussed in later groups before a final decision was made about adding or removing an item. Thus, the "new" draft version of the NOOS was consistently re-tested in focus groups until data saturation was achieved. After completing focus groups the modified version of the NOOS was field-tested.

An electronic version of the questionnaire, including the modifications based on focus group input, was created for subsequent development (field-testing). The electronic version was a precise reproduction of the paper-based layout. However, to ensure transferability, the electronic version was tested and assessed by 11 patients with NP and 3 HCPs (details available upon request).

Patients with the same eligibility criteria as above were recruited for field-testing from 15 private physiotherapist clinics and 1 spine centre in or near Copenhagen, Aarhus, Odense and Aalborg, Denmark. The baseline electronic data collection took place at the clinics via a designated tablet. One and 12 weeks after enrolment the participants were provided with a link to the questionnaire via email to enable questionnaire completion using a personal computer, tablet or smartphone. SurveyXact® (©2013 Rambøll) was used for data collection and storage. If the participant did not respond, a reminder was sent after 2, 4 and 6 days.

Item reduction was performed using data from field-testing. Items were removed if they had a mean and/or a median score of < 1 ("mild" problems), and/or more than 30% of the respondents reported having "no problems" (baseline administration) and/or if they had low reliability (i.e. a test-retest reliability intraclass correlation coefficient (ICC version 2,1) (33) less than 0.50 (18)). Finally, prior to conducting principal component analysis (PCA), the correlation matrix of the remaining items was assessed. Items a priori assigned to sub-scales based on the same underlying construct, were expected to have correlations between 0.30 and 0.80. Items outside this range were considered for removal. PCA with oblique rotation was conducted with items with factor loadings on a single factor ≤ 0.30 removed (34). The clinical importance of each item was reviewed by all members of the study group (TJ, EMR, KS, AMD) before any items were removed based on the above-mentioned criteria.

Construct validity of the subscales

After completing the item reduction process, construct validity was assessed. The initial plan was to confirm the subscale structure using confirmatory factor analysis (CFA). CFA is strongly driven by theory

Table I. *Correlation matrix containing the remaining "pain" and "physical function, daily activities" items (*n = 196*)*

Item	P1	P2	P3	P4	P5	P7	P8	P9	P10	P11	P12	P14
P1	1.000											
P2	0.288	1.000										
P3	0.150	0.343	1.000									
P4	0.116	0.322	0.238	1.000								
P5	0.315	0.434	0.123	0.231	1.000							
P7	0.266	0.261	0.020	0.282	0.523	1.000						
P8	0.224	0.323	0.064	0.249	0.391	0.445	1.000					
P9	0.197	0.361	0.140	0.341	0.453	0.396	0.500	1.000				
P10	0.333	0.449	0.188	0.297	0.480	0.440	0.513	0.628	1.000			
P11	0.140	0.370	0.166	0.423	0.479	0.444	0.524	0.431	0.584	1.000		
P12	0.228	0.404	0.240	0.443	0.446	0.399	0.451	0.564	0.629	0.617	1.000	
P14	0.282	0.388	0.230	0.375	0.388	0.374	0.603	0.476	0.569	0.485	0.487	1.000
A1	0.209	0.293	0.125	0.266	0.328	0.306	0.567	0.470	0.503	0.454	0.373	0.652
A3	0.251	0.299	0.142	0.345	0.368	0.361	0.445	0.652	0.573	0.426	0.485	0.501
A7	0.295	0.337	0.208	0.515	0.412	0.409	0.476	0.492	0.601	0.570	0.716	0.540
A8	0.174	0.341	0.222	0.499	0.360	0.281	0.418	0.471	0.503	0.575	0.580	0.490
A9	0.195	0.356	0.263	0.468	0.375	0.289	0.410	0.491	0.512	0.563	0.706	0.461

P1 = item 1, pain subscale, etc.; A1 = item 1, physical function, daily activities subscale, etc.
A full description of the items included in the final version of the Neck OutcOme Score questionnaire is given in Table VI.

and was therefore chosen since the ICF framework was used as a conceptual model. The items were initially arranged according to the subscales based on the ICF and confirmatory factor analysis was initiated. However, due to poor model fit, item relationships including within the assigned subscales based on correlations, debating conceptual fit, item correlations and possible issues of different item response structure were evaluated. Examination of the inter-item correlation matrix, e.g. (Table I) revealed that many items had correlations within the range 0.30–0.80 with items on other subscales, making it difficult to achieve a model with overall goodness of fit (34) and conceptual clarity. Hence, we elected to conduct an exploratory factor analysis (EFA) with promax rotation (with Kaiser normalization). Parallel analysis and the scree test were assessed to determine the number of factors retained (35). Items were considered to load on a factor if the factor loading was ≥ 0.30 (34). A sample size of 5–7 patients per item has been recommended when conducting factor analysis as long as there are at least 100 patients in the sample (36). All quantitative data analyses were conducted in SPSS v22.0 (IBM®, SPSS Statistics).

RESULTS

Item generation and content validity

A total of 11 focus groups (6 focus groups including 24 patients (19 women, 5 men (mean age 57.2 (standard deviation (SD) 15.9), range 24–85 years)) and 5 focus groups with HCP) were conducted before data saturation was achieved. The HCPs consisted of 6 physiotherapists, 2 professors, 3 associate professors and 1 PhD student. Finally, 1 linguist (professor) was interviewed separately to ensure comprehensibility of the questionnaire. All focus groups were conducted from January to April 2012.

As shown in Table II, 19 of 69 items were removed because they were deemed unimportant, irrelevant or redundant. Four items were added to the questionnaire because patients felt they were missing. These items were related to swallowing, hearing and balance difficulties (symptom subscale) and experience of limitations in one's life as a result of neck problems (neck-related QOL subscale). In addition, "finding joy in everyday life" was moved from the "emotional impact" subscale to "neck-related QOL" subscale as patients' perceived this an essential part of QOL. Items that patients' perceived as symptoms associated with their neck problems (e.g. headache) were moved from the "general health" subscale to the "symptoms" subscale. NP when sitting still for a long period of time was re-phrased to incorporate sitting still for more than 1-h based on patient feedback. The last response option initially worded "extreme" was changed to "very strong" (except for items related to pain intensity) as several patients mentioned that they would never choose a word like "extreme" to describe their situation.

The HCPs added 5 items to the questionnaire: headache intensity and arm/finger pain intensity ("pain" subscale); visual disorder, difficulty tying a shoelace, writing and sleep difficulties in relation to neck problems were added to the "symptom" subscale. Four items related to pain during sitting still were merged into 1 item incorporating pain during a sitting activity, such as reading, watching TV (pain subscale). NP when "standing still for a short period of time" was re-phrased as "standing still for more than 30 min" and a similar

Table II. *Initial item reduction**

Subscales	Reasons for removal of items	Total number of items removed (n=19)
Symptoms		
Do you have swelling in your neck?	Not important to the patients	1
Pain		
Turning your neck?	The items were either directly adapted from the KOOS,	4
Moving your neck?	which the neck pain patients found irrelevant or the activity	
Going for a walk?	was covered in 1 of the other items	
Going up or down stairs?		
Physical function, daily activities		
Descending stairs?	All items were adapted directly from the KOOS and were	8
Ascending stairs?	"knee-related" activities, which were deemed irrelevant or	
Bending to floor/pick up an object?	not important by the neck pain patients	
Walking on flat surface?		
Putting on socks/stockings?		
Taking off socks/stockings?		
Getting in/out of bath?		
Getting on/off toilet?		
Function in sport and recreation		
Squatting?	The items were either directly adapted from the KOOS,	5
Jumping?	which the neck pain patients found irrelevant or the activity	
Running?	was covered in one of the other items	
Keeling?		
Perform activities with your arms above your head?		
Quality of life		
How much are you troubled with lack of confidence in your neck?	Not important to the neck pain patients	1

*Preliminary items derived from patients were not linguistically validated into English since they were not included in the final questionnaire.
All items were removed after the first 2 patient focus groups.
KOOS: Knee injury and Osteoarthritis Outcome Score.

item was removed from the "activity" subscale. Furthermore, the HCPs removed 1 item each from the "neck-related QOL" and "emotional impact" subscales. Both items were linked to daily life situations taking place prior to patients experiencing neck problems and might be subject to recall bias. Finally, 2 items about the future concerns and impact were removed, as the HCPs were concerned that these items would not change during a rehabilitation intervention. As only 2 items remained in the "emotional impact" subscale, there were concerns about the psychometric properties of a subscale containing only 2 items, and these items were therefore moved to the "neck-related QOL" subscale.

The resulting 54-item draft questionnaire with 5 subscales; symptoms (18 items), pain (14 items), physical function, daily activities (9 items), participation (7 items) and neck-related QOL (6 items), was reviewed by a linguist, who made minor changes to the syntactic structure and wording of items. After pilot-testing in 2 focus groups, no further modifications were identified and this preliminary version in electronic format was used for field-testing.

Item reduction based on the electronic survey data

A total of 196 patients (146 women, mean age 47.8 (SD 13.7), range 18–89 years) with NP completed the questionnaire. Hereof 43% also reported experiencing head pain/headache, 55% reported shoulder pain (left, right or both), 24% reported pain in the upper arm (left, right or both) and 19% reported experiencing pain in the forearm/hand (left, right or both). Due to electronic data collection, there were no missing data. Patient characteristics are presented in Table III.

After review of the patient responses, content, measurement qualities and clinical relevance by all 4 authors using the a priori criteria, there was consensus that 18 items met criteria for removal. Fig. 1 and Table IV show the items removed through the quantitative item reduction processes. As shown in Table IV, 9 items were removed from the "symptom" subscale. Item S5, S11, S12 and S13 had a high proportion of patients reported having "no problems" ($\geq 51\%$); S4 had a low correlation with the remaining items (≤ 0.28) and, along with S15, did not meet factor loading criteria; S6 had a very low correlation with the remaining items (≤ 0.24), except for 1 (S8 = 0.30) and even

Table III. *Demographic characteristics of patients with neck pain*

Characteristics	Baseline ($n=196$)
Female, n (%)	146 (74.5)
Age, years, mean (SD)	47.8 (13.7) ($n=187$)
BMI, kg/m², mean (SD)	25.6 (5.1) ($n=184$)
Duration of neck pain, n (%)	
0–2 weeks	6 (3.1)
2–4 weeks	12 (6.1)
1–3 months	16 (8.2)
3–6 months	11 (5.6)
More than 6 months	151 (77)
Episodes, n (%)	
First time	41 (20.9)
1–3 episodes	33 (16.8)
4–6 episodes	13 (6.6)
7–10 episodes	3 (1.5)
More than 10 episodes	106 (54.1)
UCLA activity-level rating, n (%)	
Regularly participate in impact sports such as jogging, tennis, skiing, acrobatics, ballet, heavy labour, or backpacking	27 (13.8)
Sometimes participate in impact sports	30 (15.3)
Regularly participate in very active events, such as bowling or golf	9 (4.6)
Regularly participate in active events, such as bicycling	28 (14.3)
Regularly participate in moderate activities, such as swimming and unlimited housework or shopping	28 (14.3)
Sometimes participate in moderate activities	23 (11.7)
Regularly participate in mild activities, such as walking, limited housework, and limited shopping	17 (8.7)
Sometimes participate in mild activities	28 (14.3)
Mostly inactive: restricted to minimal activities of daily living	5 (2.6)
Wholly inactive: dependent on others; cannot leave residence	1 (0.5)
Education level, n (%)	
Primary school	22 (11.2)
High school	77 (39.3)
Undergraduate	67 (34.2)
Graduate	28 (14.3)
Did not wish to inform	2 (1.0)

BMI: body mass index; UCLA: University of California Los Angeles activity-level rating.
Ground school = 9–10 years of education; High school = 10–12 years of education; Undergraduate = 13–15 years of education; Graduate = 15–17 years of education.
SD: standard deviation.

Table IV. *Removal of items from the preliminary NOOS subscales* (quantitative criteria)*

Subscales	More than 30% of the respondents reporting having "no problems"	Mean <1	Median <1	ICC <0.50	Low correlation with the remaining items (<0.30)	High correlation with the remaining items (>0.80)	Failure to load on 1 single factor (≥0.30)
Symptoms							
S4. Have you felt a grinding sensation. heard clicking or any other noise from your neck?	–	–	–	–	×	–	×
S5. How often has your neck for example got stuck or been "locked"?	×	–	–	–	–	–	0
S6. Have you felt any numbness or a tingling sensation in your arm or fingers?	–	–	–	–	×	–	0
S10. Have you experienced balance difficulties?	×	–	–	–	×	–	0
S11. Have you experienced any swallowing difficulties?	×	×	×	–	0	0	0
S12. Have you experienced humming. buzzing in the ears or similar hearing difficulties?	×	–	×	–	0	0	0
S13. Have you experienced double vision, had difficulties focusing or similar visual disorders?	×	×	–	–	0	0	0
S15. Have you been more tired than usual?	–	–	–	×	–	–	×
S16. Overall, how stiff has your neck been?	–	–	–	×	–	–	0
Pain							
P4. How is your arm and finger pain when it is at its worst?	×	–	–	–	–	–	×
P6. Have looked down at your chest?	–	–	–	×	0	0	0
P13. Have lifted or carried something light, e.g. a milk carton. a cup, etc.?	×	×	×	–	0	0	0
Physical function, daily activities							
A2. Have been getting out of bed?	×	×	–	–	0	0	0
A4. Have had to stand up from a sitting position?	×	×	–	–	0	0	0
A5. Have had to hold an object, e.g. a cup, a glass, tableware, etc.?	×	×	×	–	0	0	0
A6. Have had to get in or out of a car?	×						
Participating in everyday life							
PT5. Have had to transport yourself to wherever you want to go, e.g. by bus, train. car, bike, etc.?	×	–	–	–	–	–	0
Quality of life							
Q6. To what extent can you control your neck problems, e.g. ease or reduce your neck problems?	–	–	–	–	×	–	0

Items were removed if they had: (*i*) a mean score of <1; (*ii*) and/or a median score of <1 (indicating "mild" problems); (*iii*) and/or more than 30% of the respondents reported having "no problems"; and/or (*iv*) a test-retest reliability coefficient less than 0.50. The correlation matrix of the remaining items was assessed with expected correlations between 0.30 and 0.80. If a given item did not correlate with any of the similar items, they were considered for removal. Finally, failure to load on 1 single factor (≥0.30) suggested that a given item did not measure the underlying construct and was therefore considered for deletion.

–=not fulfilling the criteria; ×=fulfilling the criteria; 0=not included in the analysis.

ICC: intraclass correlation coefficient; NOOS: Neck OutcOme Score. P4=item 4, pain subscale, etc.; A2=item 2, physical function, daily activities subscale, etc.; S4=item 4, symptom subscale, etc.; PT5=item 5, participating in everyday life subscale; Q6=item 6, neck-related quality of life subscale.
*Unvalidated translation into English.

though the item is clinically relevant, it was decided to remove it. All members of the study group agreed that feeling numbness or a tingling sensation in their arm or fingers represented a separate construct, which in the current questionnaire is only represented by this 1 item. S16 was highly correlated with S17 and S18 ($r=>0.74$) and it had low reliability (ICC=0.58; 95% confidence interval (95% CI): 0.44–0.72); and, S10 only correlated with S8 and S9 and 68% of the patients reported having "no problems" or "rarely" having problems. Three items were removed from the "pain" subscale: P6 had poor reliability (ICC=0.38; 95% CI 0.09–0.59); P13 had a mean and a median

<1; and, P4 had low factor loadings (<0.23) and >30% of the patients reported having "no problems". Four items were removed from the "physical function, daily activities" subscale. A2, A4, A5 and A6 were considered irrelevant since they all had mean values <1 and >30% of the patients reported having "no problems". PT5 was highly correlated with PT1, PT2, and PT3 (>0.62) and 38% of the patients reported having "no problems". The item was therefore considered irrelevant and removed from the "participation" subscale. Finally, Q6 was removed from the "neck-related QOL" subscale due to low correlations with the remaining "neck-related QOL" items (<0.30).

After this item reduction process, the questionnaire consisted of 36 items (symptoms (9); pain (11); physical function, daily activities (5); participation (6); and, neck-related QOL (5)). These items were included in the assessment of subscale structure.

Construct validity of the subscales

The EFA was conducted after the item reduction process including 36 items, and results from the final EFA (promax rotation) are presented in Table V.

Table V. *Exploratory factor analysis with rotation – pattern matrix (*n = 196)

Item	Factor 1 (everyday activity and pain)	Factor 2 (participating in everyday life)	Factor 3 (mobility)	Factor 4 (sleep disturbance)	Factor 5 (symptoms)
A9	**0.923**				
P12	**0.911**				
A7	**0.828**				
A8	**0.808**				
P11	**0.664**				
D2	0.604	**0.426**			
D6	0.573	**0.327**		−0.276	
D3	0.517	**0.361**			−0.217
D7	0.509	**0.481**		−0.280	
D1	0.493	**0.334**			
P10	**0.476**			0.352	
D4	0.476	**0.282**			
P9	**0.417**			0.409	
Q4		**0.814**			
Q2		**0.771**			
Q3		**0.757**			
Q5		**0.675**			
S1			**0.853**		−0.205
S2			**0.848**		
P7			**0.705**		
S18			**0.692**		
S3			**0.683**		
S17	−0.203		**0.620**		
P5	0.282		**0.611**		
P14				**0.807**	
S14				**0.789**	
P8				**0.770**	
A1		0.274		**0.740**	
A3	**0.330**			0.426	
P3					**0.888**
S7					**0.818**
S8					**0.636**
S9	0.242	0.257			**0.441**
P2			0.231		**0.404**
Eigenvalue	13.532	2.587	1.967	1.708	1.319
Amount of variance explained by the factor (%)	39.800	7.609	5.784	5.025	3.879
Total amount of variance explained by the factor (%)	39.800	47.409	53.193	58.218	62.096

P1=item 1, pain subscale, etc.; A1=item 1, physical function, daily activities subscale, etc.; S1=item 1, symptom subscale, etc.; PT1=item 1, participating in everyday life subscale, etc.; Q1=item 1, neck-related quality of life subscale, etc.

Amount of variance (%)=percentage of variance explained by the latent factor.

Total amount of variance (%)=cumulated percentage of variance explained by the 5 factors.

Bold equals the items included in the given subscale.

The analysis yielded 6 factors with an eigenvalue over 1 and factor loadings of 0.30 or above, explaining 64% of the variance. Parallel analysis indicated that the 6 factors should be retained and χ^2 goodness-of-fit analysis revealed that the 6-factor solution fit the data, χ^2= 4676.42, $p < 0.000$. Item Q1 and P1 separated from the remaining items, loading on a single factor with an eigenvalue of 1.38 explaining 4% of the variance. However, both items refer to frequency ("How often do you feel NP?" and "How often are you made aware of your neck problems?"), which is significantly different from the other items referring to the degree of a given problem. Furthermore, a subscale consisting of only 2 items (which may represent a method effect given their response options) often has psychometric challenges (13). The factor analysis was repeated after omitting item P1 and Q1. It did not significantly change the remaining factor loadings. Thus, the items were removed (Table VI).

After performing factor analysis the subscales were renamed, as the initial a priori subscales (latent constructs) were perceived to be somewhat altered, and not well reflected by the initial labels. As shown in Tables V and VI, the final version of the questionnaire thus consisted of 34 items divided into the following subscales: Mobility (7 items: S1, S2, S3, S17, S18, P5, P7), Symptoms (5 items: P2, P3, S7, S8, S9), Sleep disturbance (4 items: P8, P14, S14, A1), Everyday activity and pain (8 items: P9, P10, P11, P12, A3, A7, A8, A9), Participation in everyday life (10 items: PT1, PT2, PT3, PT4, PT6, PT7, Q2, Q3, Q4, Q5). A description of the items included in the final version of the NOOS questionnaire is presented in Table V.

The Danish and English linguistically translated versions (37) are available from http://www.koos.nu/.

DISCUSSION

The development of a multidimensional PRO for patients with NP, specifically item generation, reduction and construct validity, is described here.

During the qualitative development process patients with NP uncovered issues related to their neck problems, revealing complications such as movement difficulties, headache and dizziness, but more unforeseen issues related to participating in leisure activities and experiencing limitations in one's life were also revealed. Hence, the NOOS covers a broad spectrum of important aspects of life.

Content validity was achieved through the inclusion of stakeholders, such as patient and

Table VI. *Final version of NOOS items and subscales (n = 196)*

Subscales	Answer options: 0=no problems. 4=extreme problems. Frequency (%)					Mean	Median	ICC (95% CI) (n=92)
	0	1	2	3	4			
Mobility								
S1. Have you been able to turn your head fully and without difficulty?	19	20	23	14	23	2.04	2	0.69 (0.56–0.78)
S2. Have you been able to tilt your neck or head all the way back without difficulty?	22	18	21	22	17	1.95	2	0.69 (0.57–0.78)
S3. Have you been able to look down at your chest without difficulty?	8	15	16	24	37	2.68	3	0.46 (0.06–0.73)
S17. How stiff is your neck when you have just woken up in the morning?	5	28	30	29	9	2.10	2	0.69 (0.56–0.78)
S18. How stiff has your neck been later in the day?	8	26	40	20	6	1.90	2	0.58 (0.41–0.70)
P7. Have you turned your head all the way to one side?	10	31	35	22	2	1.74	2	0.55 (0.39–0.68)
P5. Have you tilted your head or neck all the way back?	13	30	36	19	2	1.67	2	0.58 (0.42–0.70)
Symptoms								
P2. What is your neck pain like when it is at its worst?	1	6	31	52	10	2.64	3	0.68 (0.55–0.78)
S7. Have you had a headache?	14	14	32	30	10	2.08	2	0.78 (0.68–0.87)
P3. What is your headache like when it is at its worst?	13	9	18	41	19	2.43	3	0.88 (0.83–0.93)
S8. Have you felt dizzy?	26	26	31	15	2	1.41	1	0.74 (0.63–0.83)
S9. Have you had trouble concentrating?	15	23	39	16	6	1.74	2	0.80 (0.72–0.87)
Sleep disturbance								
P8. Have you been lying down?	18	40	34	8	1	1.33	1	0.57 (0.41–0.69)
P14. What degree of neck pain have you felt at night for example pain that disturbed your sleep?	24	29	30	15	2	1.40	1	0.70 (0.58–0.79)
A1. Have you lain in bed?	33	28	24	13	2	1.23	1	0.59 (0.44–0.71)
S14. How often have you slept badly because of your neck problems?	12	16	33	24	15	2.14	2	0.75 (0.64–0.83)
Everyday activity and pain								
P9. Have you sat still for more than one hour when for example reading, watching TV or sitting in front of a computer?	4	23	41	30	3	2.04	2	0.68 (0.55–0.78)
P10. Have you been standing for more than 30 minutes?	18	33	40	8	1	1.39	1	0.70 (0.57–0.79)
P11. Have had your arms above your head for example when getting dressed, washing or brushing your hair?	34	28	34	11	3	1.22	1	0.70 (0.58–0.79)
P12. Have lifted or carried heavy items, such as grocery bags?	18	22	34	20	5	1.71	2	0.77 (0.67–0.84)
A3. Have sat still for more than 1 hour?	15	34	36	11	4	1.55	2	0.64 (0.50–0.74)
A7. Have been shopping?	33	36	20	8	3	1.11	1	0.75 (0.65–0.83)
A8. Have been doing light housework, such as cooking or dusting?	32	34	22	10	2	1.16	1	0.80 (0.72–0.87)
A9. Have been doing heavy housework, such as washing the floor or vacuuming?	22	26	22	20	9	1.67	2	0.81 (0.72–0.87)
Participating in everyday life								
PT1. Have been taking part in social life, such as visiting your family, friends or colleagues?	36	34	19	8	3	1.09	1	0.73 (0.60–0.80)
PT2. Have been practising your preferred leisure activities, such as hobbies or handicrafts?	23	31	26	14	6	1.48	1	0.65 (0.51–0.75)
PT3. Have been practising your preferred sporting activities, such as swimming, cycling, running or tennis?	20	33	24	16	7	1.56	1	0.56 (0.40–0.68)
PT4. Have been attending to your work or studies in or outside your home?	23	29	27	14	8	1.54	1	0.81 (0.72–0.87)
PT6. Have you been able to participate in your preferred physical activities for as long as you would like?	21	22	26	18	12	1.77	2	0.59 (0.44–0.71)
PT7. Have you been able to participate in your preferred physical activities in the manner you would like?	18	25	24	19	13	1.84	2	0.70 (0.57–0.79)
Quality of life								
Q2. Have your neck problems caused you to make changes to your lifestyle?	12	32	12	32	13	2.02	2	0.77 (0.67–0.84)
Q3. Do you experience limitations in your life as a result of your neck problems, such as avoiding or limiting work, leisure activities, hobbies or socialising?	20	29	15	20	16	1.82	2	0.78 (0.69–0.85)
Q4. Have your neck problems affected your relations to those closest to you?	48	21	13	11	7	1.07	1	0.87 (0.80–0.91)
Q5. Do your neck problems have an emotional affect on you, for example in the form of experiencing sadness, frustration or anger?	19	37	18	16	10	1.62	1	0.73 (0.60–0.81)

NOOS: Neck OutcOme Score; P1 =item 1, pain subscale, etc.; A1 =item 1, everyday activity and pain, etc.; S1 =item 1, symptom subscale, etc.; PT1 =item 1, participating in everyday life subscale; Q1 =item 1, neck-related quality of life subscale; ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval.

clinical HCPs. The development process covered the inclusion of items addressing content from existing PRO questionnaires, such as the NDI (7) and the Neck Pain and Disability Scale (29), but also inclusion of new items that emerged during focus groups, and a new subscale was formed (participation). To underpin the concepts being measured the ICF framework was used. Hence, while the identification of items pertaining to participation might have been expected, patients who were not knowledgeable about the framework, generated these new items. The emergence of new items confirms that the NOOS addresses gaps in existing PROs and stresses the need for a new measure.

The ICF framework includes 3 domains: body functions and structure, activity, and participation. In addition, it incorporates contextual factors, divided into external environmental factors (e.g. social structures) and internal personal factors (e.g. social background) (20). While the NOOS covers the 3 domains, it does not include environmental or personal factors. In the ICF, environmental factors are a single classification, while personal factors are included in the model, they are not classified (14), and hence personal factors are extremely difficult to categorize systematically. Furthermore, the aim was to develop an instrument that could evaluate rehabilitation programmes for patients with NP. It was developed in a clinical setting with patients seeking physiotherapy treatment. Personal and environmental factors are conceptually likely to be effect modifiers or may mediate the relationship between the main ICF components. As a result, the questionnaire was developed focusing on body functions and structure, activity and participation. The emergence of new items confirms that the NOOS addresses gaps in the existing PROs.

The ICF manual provides guidance regarding classification of content, but there are several aspects that have been left open awaiting further elaboration (17); for example, a clear distinction between the activity and participation domains, additional specification of the conceptual model, particularly the relationship(s) of the different components (14). From the patients' perspective, our findings show that there was no clear distinction between the body functions and structure and activity domains when patients were asked to respond to symptoms or difficulties experienced in relation to specified activities, which might explain why many of the existing measurements have failed to detect the effects of rehabilitation interventions. However, there was clear differentiation between the activity and participation domains. Lack of empirical discrimination challenges the investigation of validity. Similar results have been found for other PROs (38). As a result, researchers have argued that discriminant validity between, for example, pain on activity and self-reported physical functions measures, is lacking (38). Additional research is needed in order for the ICF framework to have both a theoretical and an empirical foundation, e.g. clearly distinguishing the boundaries between the different domains. While our data resulted in abandoning our confirmatory factor analytical approach, the results of the EFA support 5-subscale structure of the NOOS. However, it is important that further empirical testing of the NOOS incorporates CFA.

The present study included patients across the breadth of NP and recruited patients from both private therapy clinics and the municipality; we did not include people with NP who did not seek therapy. This may limit the generalizability of the results, as there may be differences between patients not seeking and seeking therapy. However, access to rehabilitation is publically available in Denmark, which might have limited potential differences.

During the focus groups participants completed the preliminary version of the questionnaire, which was ICF-based. This could have affected the dialogue to only include issues related to the constructs (items) presented, thus limiting the opportunity for breeding new constructs. Nevertheless, great effort went into facilitating an open dialogue beyond the presented items, although it cannot be ruled out that this may have influenced the outcome. Furthermore, the HCPs removed 2 items related to future concerns and impact, which may have limited the measure as it has been shown that patient's beliefs about the consequences of their condition can have an impact on rehabilitation (39).

In conclusion, the NOOS has excellent content validity, and preliminary results support a 5-subscale structure: Mobility (7), Symptoms (5), Sleep disturbance (4), Everyday activity and pain (8), and Participating in everyday life (10). Further work is needed to confirm test-retest reliability (measurement error, internal consistency), construct validity, responsiveness and interpretability. In addition, our data indicate ways in which the ICF framework could be developed.

Conflicts of interest

The authors declare that they have no financial affiliation (including research funding) or involvement with any commercial organization that has a direct financial interest in any matter included in this manuscript. The authors declare that they have no conflicts of interest.

Authors' contributions

Tina Juul was involved in the planning of the study design, data acquisition, data analysis and writing the paper. Ewa M Roos, Karen Søgaard and Aileen Davis contributed to the analysis and interpretation of the data as well as study conception and design. All authors were involved in drafting the article or revising it critically for important intellectual content and all authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

REFERENCES

1. Fernández-de-Las-Peñas C, Alonso-Blanco C, Hernández-Barrera V, Palacios-Ceña D, Jiménez-García R, Carrasco-Garrido P. Has the prevalence of neck pain and low back pain changed over the last 5 years? A population-based national study in Spain. Spine J

2013; 13: 1069–1076.

2. Hogg-Johnson S, van der Velde G, Carroll LJ, Holm LW, Cassidy JD, Guzman J, et al. The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. Spine (Phila Pa 1976) 2008; 33: 39–51.

3. Sjolander P, Michaelson P, Jaric S, Djupsjobacka M. Sensorimotor disturbances in chronic neck pain, range of motion, peak velocity, smoothness of movement, and repositioning acuity. Man Ther 2008; 13: 122–131.

4. Bexander CS, Mellor R, Hodges PW. Effect of gaze direction on neck muscle activity during cervical rotation. Exp Brain Res 2005; 167: 422–432.

5. Falla D, O'Leary S, Farina D, Jull G. The change in deep cervical flexor activity after training is associated with the degree of pain reduction in patients with chronic neck pain. Clin J Pain 2012; 28: 628–634.

6. Holtermann A, Hansen JV, Burr H, Søgaard K. Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. S Scand J Work Environ Health 2010; 36: 34–41.

7. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 1991; 14: 409–415.

8. Ailliet LK, Rubinstein DL, Sidney M, De Vet HCW, Van Tulder MW, Terwee CB. Definition of the construct to be measured is a prerequisite for the assessment of validity. the Neck Disability Index as an example. J Clin Epidemiol 2013; 66: 775–782.

9. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res 2012; 21: 659–670.

10. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 2009; 39: 400–417.

11. Vernon H. The Neck Disability Index: state-of-the-art, 1991–2008. J Manipulative Physiol Ther 2008; 31: 491–502.

12. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Heal 2007; 10: 125–137.

13. De Vet HC, Terwee CB, Mokkinik LB, Knol DL. Measurement in medicine a practical guide. First. New York, NY 10012: Cambridge University Press; 2011.

14. World Health Organization (WHO). International Classification of Functioning, Disability and Health: ICF. Geneva: WHO; 2001. Available from: http://www.disabilitaincifre.it/documenti/ICF_18.pdf.

15. Tate RL, Godbee K, Sigmundsdottir L. A systematic review of assessment tools for adults used in traumatic brain injury research and their relationship to the ICF. Neuro Rehabilitation 2013; 32: 729–750.

16. Franki I, Desloovere K, De Cat J, Feys H, Molenaers G, Calders P, et al. The evidence-base for conceptual approaches and additional therapies targeting lower limb function in children with cerebral palsy: a systematic review using the ICF as a framework. J Rehabil Med 2012; 44: 396–405.

17. Badley EM. Enhancing the conceptual clarity of the activity and participation components of the International Classification of Functioning, Disability, and Health. Social Sci Med 2008; 66: 2335–2345.

18. Thorborg K, Hölmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. Br J Sports Med 2011; 45: 478–491.

19. Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. Psychol Methods 2013; 5: 155–174.

20. World Health Organization (WHO). Towards a common language for functioning, disability and health – ICF. Geneva: World Health Organization; 2002, p. 1–22.

21. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS) – development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28: 88–96.

22. Nilsdotter AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) - validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003; 4: 10.

23. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. Foot Ankle Int 2001; 22: 788–794.

24. Bolton JE, Humphreys BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. J Manipulative Physiol Ther 2002; 25: 141–148.

25. Jordan A, Manniche C, Mosdal C, Hindsberger C. The Copenhagen Neck Functional Disability Scale: a study of reliability and validity. J Manipulative Physiol Ther 1998; 21: 520–527.

26. Leak AM, Cooper J, Dyer S, Williams KA, Turner-Stokes L, Frank AO. The Northwick Park Neck Pain Questionnaire, devised to measure neck pain and disability. Br J Rheumatol 1994; 33: 469–474.

27. Pinfold M, Niere KR, O'Leary EF, Hoving JL, Green S, Buchbinder R. Validity and internal consistency of a whiplash-specific disability measure. Spine (Phila Pa 1976) 2004; 29: 263–268.

28. Rebbeck TJ, Refshauge KM, Maher CG, Stewart M. Evaluation of the core outcome measure in whiplash. Spine (Phila Pa 1976) 2007; 32: 696–702.

29. Wheeler AH, Goolkasian P, Baird AC, Darden BV. Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. Spine (Phila Pa 1976) 1999; 24: 1290–1294.

30. White P, Lewith G, Prescott P. The core outcomes for neck pain: validation of a new outcome measure. Spine (Phila Pa 1976) 2004; 29: 1923–1930.

31. Terwee CB, Schellingerhout JM, Verhagen AP, Koes BW, de Vet HCW. Methodological quality of studies on the measurement properties of neck pain and disability questionnaires: a systematic review. J Manipulative Physiol Ther 2011; 34: 261–272.

32. Nordin M, Carragee EJ, Hogg-Johnson S, Weiner SS, Hurwitz EL, Peloso PM, et al. Assessment of neck pain and its associated disorders: results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. Eur Spine J 2009; 18: 435.

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

34. Brown TA. Confirmatory factor analysis for applied research. First. New York, NY 10012: The Guilford Press; 2006.

35. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. Organizat Res Methods 2004; 7: 191–205.

36. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. Psycholog Assess 1995; 7: 286–299.

37. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the Cross-Cultural Adaptation of Health Status Measures, New York; 2002: 1–36.

38. Pua Y-H, Cowan SM, Wrigley TV, Bennell KL. Discriminant validity of the Western Ontario and McMaster Universities Osteoarthritis index physical functioning subscale in community samples with hip osteoarthritis. Arch Phys Med Rehabil 2009; 90: 1772–1777.

39. Wertli MM, Rasmussen-Barr E, Held U, Weiser S, Bachmann LM, Brunner F. Fear-avoidance beliefs-a moderator of treatment efficacy in patients with low back pain: a systematic review. Spine J 2014; 14: 2658–2678.