University of Birmingham
Research at Birmingham

Comparison between two patient-reported outcome measures for patients with cervical radiculopathy

Thoomes, Erik; de Graaf, Marloes; Gallina, Alessio; Falla, Deborah; Stathi, Afroditi

DOI:
10.1016/j.msksp.2023.102764

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Thoomes, E, de Graaf, M, Gallina, A, Falla, D & Stathi, A 2023, 'Comparison between two patient-reported outcome measures for patients with cervical radiculopathy: A think-aloud study', *Musculoskeletal Science and Practice*, vol. 65, 102764. https://doi.org/10.1016/j.msksp.2023.102764

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

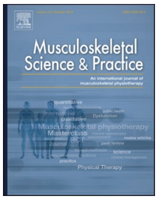
Download date: 28. Apr. 2024

Musculoskeletal Science and Practice 65 (2023) 102764

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp



Original article

Comparison between two patient-reported outcome measures for patients with cervical radiculopathy: A think-aloud study

Erik Thoomes [a,b,*], Marloes de Graaf [b,c], Alessio Gallina [a], Deborah Falla [a], Afroditi Stathi [d]

[a] Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom
[b] Fysio-Experts, Research Department, Hazerswoude, the Netherlands
[c] Department of Manual Therapy, Breederode University of Applied Science, Rotterdam, the Netherlands
[d] School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom

ARTICLE INFO

Keywords:
Cervical Radiculopathy Impact Scale (CRIS)
Patient Specific Functional Scale 2.0 (PSFS 2.0)
Patient reported outcome measure (PROM)
Cervical radiculopathy (CR)

ABSTRACT

Background: The Cervical Radiculopathy Impact Scale (CRIS) and Patient Specific Functional Scale 2.0 (PSFS 2.0) are patient-reported outcome measures (PROMs) used to assess activity limitations in patients with cervical radiculopathy (CR). This study a) compared the CRIS subscale 3 and the PSFS 2.0 in patients with CR with respect to completeness and patient preference, b) established the correlation between both PROMs in assessing the individual patient's level of functional limitations and c) assessed the frequency of reported functional limitations.

Methods: Participants with CR participated in semi-structured, individual, face-to-face interviews as part of a "think-aloud" process; verbalising their thoughts while completing both PROMs. Sessions were digitally recorded and transcribed verbatim for analysis.

Results: Twenty-two patients were recruited. The most frequently reported functional limitation on the CRIS was: 'working at a computer' (n = 17) and overhead activities' (n = 10) for the PSFS 2.0. There was significant moderate positive correlation between the scores on the PSFS 2.0 and the CRIS (Spearman's $r = 0.55$, $n = 22$ $p = .008$). Most patients (n = 18; 82%) preferred the ability to present their own individual functional limitations of the PSFS 2.0. Eleven participants (50%) preferred the 11-point scale of the PSFS 2.0 over the 5-point Likert scale scoring option of the CRIS.

Conclusion: Both easy to complete PROMs capture functional limitations in patients with CR. Most patients prefer the PSFS 2.0 over the CRIS. The wording and layout of both PROMs need refinement to enhance user-friendliness and avoid misinterpretation.

1. Introduction

Cervical radiculopathy (CR) is a clinical condition whereby motor, reflex and/or sensory changes such as paraesthesia or numbness can be present (Kuijper et al., 2009; Thoomes et al., 2012). Patients with CR can experience pain, motor weakness, sensory deficits and loss of function in the neck, shoulder, upper arm or forearm (Thoomes et al., 2012, 2021; Sleijser-Koehorst et al., 2020).

Patient Reported Outcome Measures (PROMs) assess a patient's experience of their symptoms, their functional status and their health-related quality of life (Kendrick et al., 2016). Thus, PROMs can help to determine the outcome of care in terms from the patient's perspective as an expert in the lived experience of their own health. PROMs are often self-report measures and should therefore be free of observer rating bias, but they can also be interview-based measures that involve the interviewer interpreting the patient's responses to questions (Kendrick et al., 2016). In clinical practice, PROMs are most often used to determine progress (outcomes) of individual patients (Jette et al., 2009).

The Neck Disability Index (NDI) (Vernon and Mior 1991) is often used as a PROM in studies of patients with CR (Ayub et al., 2019; Liang et al., 2019; Peolsson et al., 2019; Peolsson et al. 2019; Vleggeert-Lankamp et al., 2019; Hassan et al., 2020; Mjåset et al., 2020; Wu et al., 2020; Wu et al. 2020, Xu et al., 2020; Zhang et al., 2020). However, as patients with CR generally experience greater arm pain and periscapular

---

* Corresponding author. Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom.
E-mail address: erikthoomes@gmail.com (E. Thoomes).

https://doi.org/10.1016/j.msksp.2023.102764
Received 9 February 2023; Received in revised form 31 March 2023; Accepted 13 April 2023

pain than neck pain (Wainner et al., 2003; Kuijper et al., 2009; Thoomes et al., 2012), the NDI might not be the most suitable PROM for patients with CR (Ailliet et al., 2013; Gartner et al., 2019). A recent study reported that patients with CR rated their level of self-assessed disability on the NDI differently compared with their level of disability when assessed on the Patient Specific Functional Scale (PSFS) (Thoomes et al., 2021). NDI scores varied from 10 to 56/100 whereas the PSFS scores varied from 6.8 to 8.3/10, indicating that the NDI did not adequately assess perceived disability due to CR (Thoomes et al., 2021). Thus, the PSFS has been suggested as a more suitable PROM for patients with CR (Cleland et al., 2006). It has established reliability, construct validity and responsiveness in primary care patients with CR (Cleland et al., 2006; Horn et al., 2012; Thoomes-de Graaf et al., 2019).

The PSFS-2.0 has slightly different response options than the PSFS which requests patients to compare their current level of difficulty in performing a task to a period prior to having complaints (Thoomes-de Graaf et al., 2019). The PSFS-2.0 uses a 11-point numerical rating scale (NRS) which assesses the current self-perceived level of disability in performing a task, with '0' being 'no effort at all' and '10' being 'unable to perform'. In contrast to the PSFS (where patients define their own activity limitations), recent research suggests that the use of adding a pre-defined list of examples of activities has additional value for patients in specifying their own three most relevant activities they feel they are most restricted in (Thoomes-de Graaf et al., 2019). Patients also reported that the PSFS-2.0 was easier to complete than the PSFS (Thoomes-de Graaf et al., 2019).

In addition to the use of PSFS-2.0 as a suitable PROM for patients with CR, a new PROM was recently developed: the Cervical Radiculopathy Impact Scale (CRIS) (Gartner et al., 2019). The CRIS is a 21-item questionnaire consisting of three subscales: 'Symptoms' (9 items), 'Energy and postures' (6 items), and 'Actions and activities' (6 items). This PROM has good psychometric properties (Gartner et al., 2019). The CRIS subscale 3 asks patients to assess their perceived ability to perform six predetermined activities, whereas the PSFS 2.0 allows the patient to choose their own. However, not all predetermined activities of the CRIS subscale 3 might be relevant for each individual.

This study aims to a) compare the CRIS (subscale 3; 'actions and activities') and the PSFS 2.0 in patients with CR with respect to completeness (according to patients) and patient preference, b) establish the correlation between the PSFS-2.0 and the CRIS subscale 3 in assessing the individual patient's level of functional limitations and c) assess how often patients report certain functional limitations.

2. Methods

2.1. Design

The study involved semi-structured individual face-to-face interviews as part of a "think-aloud" process. Thinking aloud is the concurrent verbalization of thoughts while performing a task, in this case competing two PROMs. Think-aloud studies provide verbal data about reasoning during a cognitive task. Using think-aloud methodology and subsequent qualitative analysis, investigators can identify the information that participants focus on during a cognitive task and how they use that information to facilitate problem resolution. From this, inferences can be made about the reasoning processes that were used during a cognitive task, such as the completion of PROMs (Fonteyn et al., 1993, Goransson et al., 2006; Lundgren-Laine and Salantera, 2010). Recently think-aloud studies have increasingly been used to assess different aspects of clinical questionnaires, e.g., how individuals think and reason when reporting and valuing their own current health (Ernstsson et al., 2020) and the usability and feasibility of questionnaires (Kabboord et al., 2019; Jordan et al., 2020). A think-aloud qualitative assessment of the PSFS 2.0 and the CRIS subscale 3 could provide additional insight in the validity of these PROMs.

Ethical approval was granted from the University of Birmingham ethics committee (ERN_21-0891). The study adheres to the Consolidated criteria for reporting qualitative research (COREQ) criteria (Tong et al., 2007).

2.2. Participants

Patients with clinical signs of CR, including signs and symptoms of radiculopathy (either motor, reflex and/or sensory changes such as paraesthesia or numbness) and evidence of radicular pain (radiating pain in the arm and/or peri-scapular region), were included. Participants were screened in the order in which they registered themselves to the study and were able to make an appointment to be interviewed. Radicular symptoms were confirmed by one of the investigators (MdG) if the Spurling's test and at least one Upper Limb Neurodynamic test (either of the median, ulnar or radial nerve) reproduced radicular pain or patient specific symptoms (Wainner et al., 2003; Thoomes et al., 2018). The CR diagnosis was confirmed by diagnostic imaging through MRI or CT-myelography (Wainner et al., 2003; Kuijper et al., 2011; Thoomes et al., 2018). Participants were excluded in case of a history of cervical surgery or co-morbidities in the upper quadrant which could limit daily activities.

In general, patients referred to and managed in secondary health care facilities have worse functional status than people managed in primary health care settings. Therefore, participants were recruited from both secondary (specialised university neurology/neurosurgery hospital departments in the Netherlands) and primary (General Practitioner clinics and direct access physiotherapy outpatient clinic in the Netherlands) health care facilities to ensure that the participants had demographic characteristics (i.e., age, sex, level of perceived disability) similar to the population in which the CRIS subscale 3 and the PSFS 2.0 scales are used. Two of the investigators (ET and MdG) work as clinicians in primary health care facilities.

According to the theoretical underpinning of the sampling methodology used in our study, participants were then further identified through a single stage purposeful sampling technique known as criterion of inclusion or 'criterion-i sampling' (Tashakkori and Teddlie 2010; Palinkas et al., 2015). Criterion-i sampling selects individuals based on the assumption that they possess knowledge and experience with the phenomenon of interest (i.e., CR) and thus will be able to provide information that is both detailed (depth) and generalizable (breadth). Participants are to some extent assumed to be representative of the group, in this case patients with CR. From the perspective of qualitative methodology, participants who meet or exceed a specific criterion or criteria possess greater knowledge of the phenomenon of interest by virtue of their experience, making them information-rich cases (Patton 2014; Palinkas et al., 2015).

It has been suggested in two studies of using the thinking aloud method for user interface testing, that after five test subjects, 77–85% of the solutions had been found (Nielsen 1994). However, previous think-aloud studies used a recruitment matrix and included at least 10 cases in each of the two recruitment matrix cells (primary and secondary health care facilities) for a total of 20 participants (M Granger Morgan et al., 2005, Guest et al., 2006; Coenen et al., 2012; Gardner and Tang 2014; Namey et al., 2016; Zacher et al., 2022). Therefore, we aimed to recruit a minimum of 20 participants.

There was no previous established relationship between the interviewer (ET) and participants. Written informed consent was obtained prior to the interviews.

2.3. Interview procedure and materials

Participants completed the Dutch version of both the CRIS subscale 3 (items no. 15–20, **APPENDIX 1**) and the PSFS 2.0 (**APPENDIX 2**) during the interview.

The complete CRIS (subscales 1, 2 and 3) shows good content validity, test-retest reliability, construct validity and is able to discriminate between groups (Gartner et al., 2019). A recent systematic review on the psychometric properties of the original PSFS mentioned the PSFS was reported to be valid, reliable, and responsive in populations with neck dysfunction (Horn et al., 2012). The PSFS 2.0 has good content validity (Thoomes-de Graaf et al., 2019). Other psychometric qualities are currently being studied.

The order in which they were completed was decided by flipping a coin. A pre-defined list of 30 examples of activities for the PSFS 2.0 was compiled of items recorded in a recent clinical study (Thoomes et al., 2021) as well as items from Dutch versions of questionnaires assessing disabilities from the neck (NDI) (Vernon and Mior 1991; Ailliet et al., 2015) and the shoulder and Pain Disability Index (SPADI) (Roach et al.,

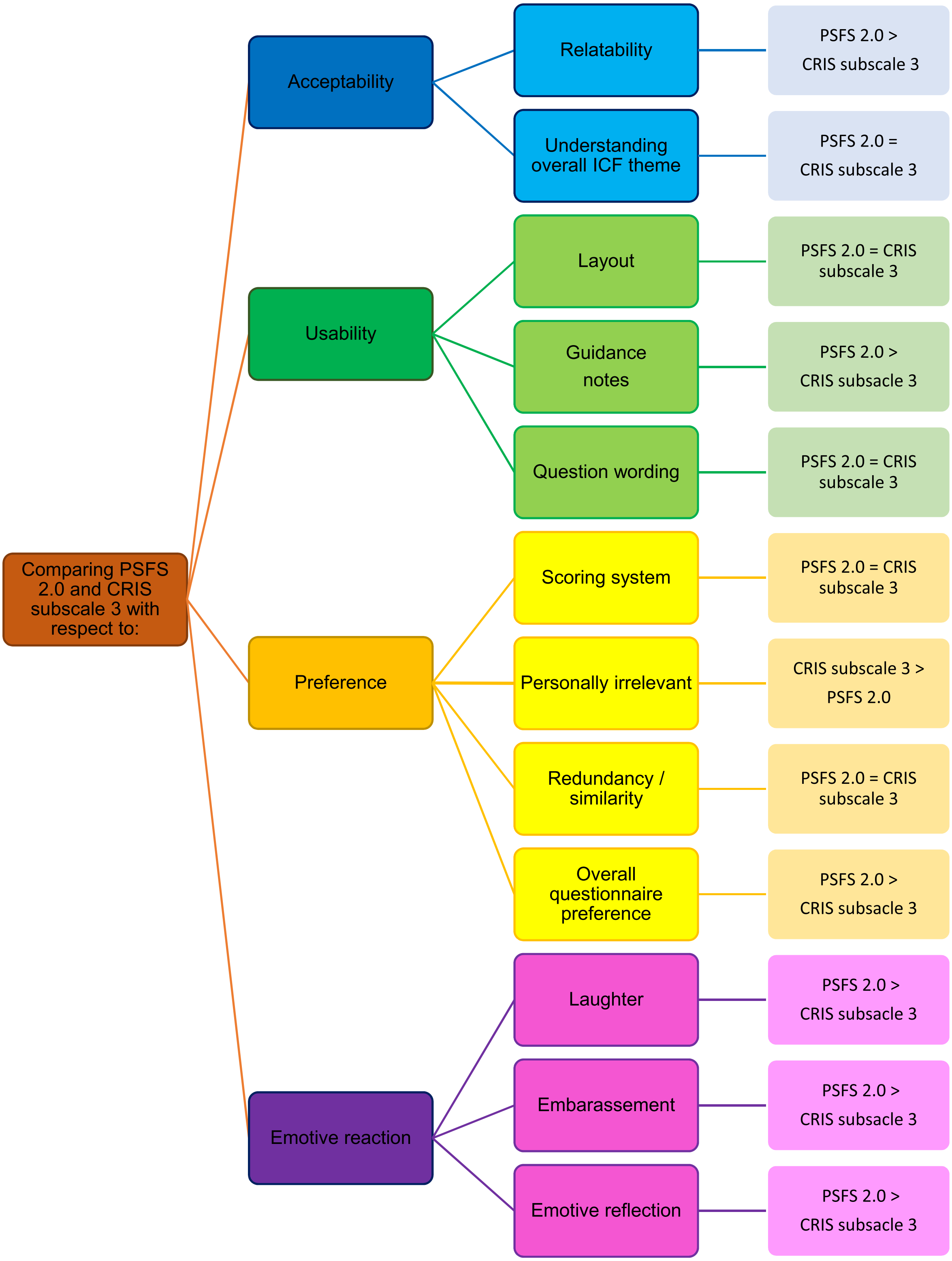


Fig. 1. Thematic tree for coding of key themes and preference of patients of the various subthemes. NB “PSFS 2.0 > CRIS subscale 3” denotes a majority preference of the PSFS 2.0; “PSFS 2.0 = CRIS subscale 3” denotes no clear preference for either PROM.

1991; Thoomes-de Graaf et al., 2017); Disabilities of the Arm Shoulder and Hand, (DASH) (Hudak et al., 1996; Veehof et al., 2002).

One male investigator (ET) conducted all face-to-face interviews in a physiotherapy clinic. Three interviews were used as a pilot, leading to minor refinements in rephrasing prompts if a participant fell silent for more than 5–7 s. The investigator was not the therapist treating the participating patients. Interviews followed a semi-structured guide (**APPENDIX 3**) and used a range of prompts as suggested in previously published think-aloud study protocols (Fonteyn et al., 1993; Pool et al., 2010) and drawing on the experience from other investigators (AS) and interviews conducted in a relevant recent think-aloud qualitative study (Thoomes-de Graaf et al., 2019). Participants were informed that the focus of the study was on their thoughts while they were answering both questionnaires rather than on the actual answers. They were instructed to think aloud: to say whatever came into their mind while responding to both questionnaires. They were informed that they might be reminded to keep talking, and that questions about items would not be answered by the investigator. If participants remained silent for 5–7 s when completing the CRIS subscale 3 or PSFS 2.0, they were asked what they were thinking about. In case verbalized thoughts were unclear, participants were asked to explain more explicitly. On completion of both questionnaires, participants were asked what they thought of completing the CRIS subscale 3 and PSFS 2.0 and if they thought something was missing in either (Fonteyn et al., 1993; Nehlin et al., 2018). They were also asked to comment on the personal relevance of activities in the CRIS subscale 3 and PSFS 2.0 and on the completeness of the example list of activities. All sessions were digitally recorded and transcribed verbatim for analysis. Participants then received the transcripts and were asked to judge if, according to their recollection, this was a true report of the interview. All data as well as the key to the participants' identifiers was stored offline on a password encrypted computer in a locked office with access only available to the researcher.

2.4. Data analysis

All statistical analyses were performed with SPSS version 28 (Inc, Chicago, Ill, USA). All data was checked for normality, using a Stem-and-leaf Plot, Q-Plot and Whisker box. Descriptive statistics were used to calculate frequencies. Spearman's rho was used to assess the correlation between outcomes on the PSFS-2.0, and CRIS subscale 3, with values of 0.0–0.3 representing "poor"; 0.3–0.5 "fair"; 0.6–0.7 "moderate"; 0.8–0.9 "very strong" and 1.0 "perfect" correlation (Chan 2003).

To analyse the think-aloud data, thematic analysis through the Framework Method was used (Gale et al., 2013; Ritchie et al., 2013; Henry et al., 2022). Themes were identified a priori by the research team while the assessors (ET and MdG) looked for additional themes as the interviews progressed. Themes were grouped in: "acceptability" (e.g., relatability, understanding of overall theme as related to activity limitations as mentioned in the International Classification of Functioning, Disability and Health), "emotive reaction" to wording (e.g., laughter, embarrassment, disappointment), "usability" (e.g., layout, guidance notes, question wording) and "preference" (e.g., redundant or similar questions and activities, personally irrelevant questions or activities, questionnaire preference, scoring system preference) (see Fig. 1).

The transcripts were read several times by two authors (ET, MdG) independently. Meaning units (words and sentences of interest for the aims of the study) were coded and sorted into key concepts after a joint discussion. Key concepts were then used to create a coding framework which could be used to code the data. Next, the thematic framework was systematically applied to code all transcripts and any coding inconsistences were discussed. The framework could be amended if new codes emerged. Coded transcripts were synthesized into a set of thematic matrix charts. The matrix charts allowed for the refinement of the overarching themes and identification of relationships, similarities, and differences within the data.

Discussion notes and log files were stored and saved automatically in the same software program used for analysing the qualitative data (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12, 2018).

2.5. Trustworthiness, Rigour and transparency

The lead researcher (ET) familiarised himself with the context of the study via conversations with participants as well as with colleagues with experience in think aloud studies and also through observations during his own clinical work with patients with CR prior to the interviews. Paraphrasing participant responses to ensure correct interpretations during the interviews was used to increase credibility and minimise bias (Noble and Smith 2015). Rigour was enhanced through a) researcher reflection on interview notes and emerging themes; b) the development of a transcription protocol; c) the employment of multiple coders. Transparency was ensured via a detailed audit trail and extensive discussion of emerging themes between coders. The COREQ checklist was used to ensure the study is reported comprehensively (**APPENDIX 4**).

3. Results

3.1. Participant characteristics

Between January and December of 2022, 22 patients (mean age 51.7 years; range 30–76 years) were recruited; 11 through referral from a neurology department (which they visited for consideration if surgery was a treatment option) and 11 through either direct access (n = 7) or referral from a local General Practitioner's clinic (n = 4) (see Table 1).

3.2. Quantitative results

The mean scores on the PSFS 2.0 and CRIS-subscale 3 were 7.6 (range 5–9.3) and 39.3 (range: 8.3–82.5) respectively. There was a moderate positive correlation between the scores on the PSFS 2.0 and the CRIS subscale 3 (Spearman's $r = 0.55$, n = 22), and the relationship was significant ($p = .008$). Most patients (n = 18; 82%) preferred the ability to present their own individual functional limitations combined with completeness of the example list of the PSFS 2.0 as a questionnaire and half (n = 11; 50%) preferred the 11-point NRS of the PSFS 2.0 over the 5-point Likert scale scoring option of the CRIS subscale 3 (see Table 2).

3.3. Functional limitations

The most often reported functional limitations in the CRIS subscale 3 i.e., those scored as either 'moderately limited', 'severely limited' or 'nearly impossible', were: 'working at a computer' (n = 17) and 'holding a book or a newspaper' (n = 12) (Table 3).

When completing the PSFS 2.0, participants most often initially reported "overhead activities' (n = 10), 'sleeping' (n = 9) and 'office/computer work' (n = 8) as being functionally limited in (Table 4).

We also merged all the activities participants mentioned from the

Table 1
Patient characteristics.

Included patients (n)	22
Age (mean, range years)	51.7; 30 - 76
Female gender (n; %)	14; 64%
Hospital referral (n; %)	11; 50%
Direct Access (n; %)	7; 32%
General Practitioner referral (n; %)	4; 18%
Mean NPRS most painful area, 0–10 (SD; range)	7.7 (1.3; 3–9)
Mean NDI score, 0–100 (SD; range)	31 (12.3; 10–56)
Area of most pain, n (%)	
Periscapular	15 (75%)
Upper arm	4 (20%)
Neck	1 (5%)
Mean duration in weeks (range)	20.7 (3–104)

Table 2

PSFS 2.0 and CRIS subscale 3 scores, preference and correlation.

PSFS score: mean; *SD* (range)	7.6; *1.2* (5.0–9.3)
CRIS subscale 3 score: mean; *SD* (range)	39.3; *19.5* (8.3–82.5)
Preferred PSFS as questionnaire (n; %)	18; 82%
Preferred PSFS scoring (Likert scale) (n; %)	11; 50%
Correlation between level of disability on PSFS 2.0 - CRIS subscale 3	$r = 0.55$, $p = .008$

Table 3

Frequency of reported limitations in the CRIS subscale 3.

Functional limitations on the CRIS subscale 3	Frequency [a] (%)
Working at a computer	n = 17 (77)
Holding a book or a newspaper	n = 12 (55)
Opening a jar with a screw-top lid	n = 8 (36)
Cutting using a knife, e.g., vegetables or meat	n = 5 (23)
Holding things in your hands without dropping them	n = 4 (18)
Writing using a pen	n = 2 (9)

[a] reported as 'moderately limited', 'severely limited' or 'nearly impossible'.

Table 4

Functional limitations most often reported in the PSFS 2.0.

Most often reported functional limitations on the PSFS 2.0	Frequency [a] (%)
Overhead activities	n = 10 (46)
Sleeping	n = 9 (41)
Office/computer work	n = 8 (36)
Driving	n = 7 (32)
Household activities	n = 6 (27)
Looking up	n = 4 (18)
Cycling	n = 4 (18)
Writing	n = 4 (18)

[a] reported as 'moderately limited', 'severely limited' or 'nearly impossible'.

PSFS and the CRIS subscale 3 and re-coded these into 14 main themes (e.g., overhead activities, carrying, driving, sleeping, office work, household duties, etc). Overall, the most frequently mentioned were 'overhead activities' (n = 37), 'household activities' (n = 35), 'office/computer work' (n = 32) and 'sleeping' (n = 30) (Table 5).

3.4. *Qualitative analysis*

Interviews were transcribed immediately after the session and data were subsequently analysed. Although themes had been identified a priori and assessors (ET and MdG) searched for additional emerging themes as the interviews progressed, no new themes were added (see Fig. 1). Forward and backward translation checks of the presented quotes were performed by two bi-lingual researchers (ET and MdG).

Table 5

Reported number of times of 14 recoded most limited activities.

Recoded activity	Number of times reported limited
Overhead activities	37
Household activities	35
Office/computer work	32
Sleeping	30
Grip-strength	14
Carrying	13
Turning	13
Reading	13
Dressing	12
Recreational activities	11
Cycling	10
Driving	9
Sexual activities	8
Looking-up	7

3.4.1. *Acceptability*

All participants understood the aim of both questionnaires (describing and rating/grading the daily activities they feel restricted in due to their CR) and the usefulness for clinicians. With respect to "relatability", most patients (n = 18) preferred the PSFS 2.0 over the CRIS subscale 3 (Fig. 1)

> "I can see how my answers can help you understand what I cannot do because of my neck and arm pain". (Female, age 44)

3.4.2. *Usability*

While completing the CRIS subscale 3, five participants initially missed completing the third question and only realized this while considering their rating on the fourth question.

> "… opening a jar … oh! I see I've missed a question about cutting vegetables or meat. Well I'll just answer this one first …" (Male, age 45)

Additionally, some participants (n = 8) initially did not see the 11-point scoring line of the PSFS 2.0 as an example. Only when reading on and seeing where they were intended to complete their three most limited activities and rate them on the 11-point NRS did they then realize this.

> "… so where do I write this (the activity they felt most limited in) down? Oh no, not here? Oh, I see now; this is the example scoring line, below on the dotted line is where I should write it down …" (Female, age 57)

This was similar to the participant's response with respect to that on the clarity of the guidance notes. When asked afterwards, none of the participants had any comments on the clarity of the guidance notes, yet several (n = 7) voiced queries with respect to completing the questions throughout completing both questionnaires.

> "… Well, in the beginning, two months ago, I was completely unable to work on a computer but this past week (interviewer interrupts and points out the CRIS subscale 3 guidance notes asks about limitations in the past week) …" (Female, age 52)

A few participants (n = 3) mentioned needing to read the guidance notes of the CRIS subscale 3 two or three times in order to make sure they understood the instructions well enough before starting to complete the questionnaire.Some of the participants commented on, or appeared to have difficulty with the question wording of the questionnaires (PSFS 2.0, n = 2 and CRIS subscale 3, n = 4).

> "… holding a book or a newspaper … I read a digital newspaper on my tablet which I hold in my other (not afflicted) hand and I use a very light E-reader as a book. But if I were to hold a traditional newspaper in the air, say in the train, that would probably cause pain. So do I answer to that imagined situation or to my own personal situation … ?" (Male, age 63)

3.4.3. *Preference*

In all cases when completing the PSFS 2.0, participants noted that some of the activities mentioned in the example list of the PSFS 2.0 were similar to activities they had already mentioned themselves in their personal three most limited activities.There was a wide variety in the number and type of questions or activities not relevant for the person mentioned in either the CRIS subscale 3 or the PSFS 2.0. This was to be expected due to the wide variety in personal circumstances of participants. Most participants (n = 18, 80%) said they preferred the PSFS 2.0 over the CRIS subscale 3, as they felt it was more tailored to them individually or personalised and they could more completely describe which specific activities they were limited in due to their CR.

> "… some of the activities mentioned in the CRIS subscale 3 I do with my other (not affected arm, so it seems I am hardly limited when I complete this …" (Male, age 45)

> "… the example list of the PSFS 2.0 allows me to more clearly list my individual limitations …" (Male, age 52)

> "… I had some difficulty in realizing which three activities I was limited in, but looking over all the examples in the PSFS 2.0 I quickly realized quite a few I was also limited in …" (Female, age 38)

Some of the questions elicited an emotive response in some of the participants. Three mentioned feeling "negatively and emotionally confronted" with their level of disability after having scrolled through the list of suggested activities in the PSFS 2.0 or having had to report having difficulty with a substantial number of them due to their CR.

> "I'm sorry, can I have a minute? I feel sort of emotional as it suddenly dawns on my how limited I am in my normal daily activities …" (Female, age 30)

Other emotional responses included being either positively or negatively surprised by one or more of the suggested activities

> "… engaging in sexual activities? No way, either through the pain but perhaps also as a side effect of the medication I do not feel any need for intimacy with my partner right now …" (Female, age 46)

> "… washing my hair? Hahaha, as you can see I am bald already so, no. I am not limited with respect to that …" (Male, age 76)

The CRIS subscale 3 elicited no clear emotive response.

4. Discussion

This is the first study to compare the PSFS 2.0 and the CRIS subscale 3 with regards to patient preference, time-to-complete, and functional limitation assessment. It is also the first qualitative think-aloud study assessing the CRIS subscale 3 and PSFS 2.0. Both the PSFS and the CRIS subscale 3 have previously shown to be able to capture the level of functional limitations in patients with CR (Cleland et al., 2006; Gartner et al., 2019). Although the NDI is commonly used for measuring neck-related disability in patients with neck pain, it cannot be used to assess upper extremity disability, especially as patients with CR usually report having more arm pain than neck pain (Wainner et al., 2003; Kuijper et al., 2009; Mehta et al., 2010; Thoomes et al., 2012). Therefore, other PROMs might be more suited for patients with CR. It is of relevance that, after recoding the individually mentioned activity restrictions into encompassing themes, the functional activities participants most often reported being restricted in were mostly upper extremity related activities. This supports the notion that the NDI might not be the best PROM for patients with CR (Ailliet et al., 2013; Gartner et al., 2019).

Developing new PROMs is a lengthy and complex process and has been advised against, especially in the light of recent developments in item response theory and computer technology (Rothrock et al., 2011). These recent developments support using the Patient-Reported Outcomes Measurement Information System (PROMIS)® instead of more traditional PROMs like the NDI (Alonso et al., 2013; Witter 2016). The PROMIS research initiative has allowed for the creation and validation of item banks for specific domains. An item bank is a collection of items assessing a single underlying trait (e.g., fatigue), with each item representing a point on the trait continuum. Item banks can be administered in multiple ways, including as a Computerized Adaptive Test (CAT) or through questionnaire forms of varying lengths (Rothrock et al., 2011). A recent study aimed to validate the association of the PROMIS with the NDI in patients with neck pain and their ability to capture concomitant arm pain and concomitant back pain (Moses et al., 2019). Clinical use of the PROMIS® item bank is still limited however, due to the financial constraints of using it.

The majority of participants in our study felt that the PSFS 2.0 was best suited to comprehensively list all functional disabilities they encountered due to their CR. In the International Classification of Functioning, Disability and Health, these are known as "participation restrictions" (Rauch et al., 2008). This list can then facilitate clinicians in breaking down 'participation restrictions' into 'activity limitations' and these then to 'body functions'. These can then be used to formulate a conservative management plan (Rauch et al., 2008). A recent study assessing the diagnostic accuracy of patient interview items, reported that 'arm pain worse than neck pain', 'provocation of symptoms when ironing', 'reduction of symptoms by walking with your hand in your pocket' showed high specificity (Sleijser-Koehorst et al., 2021). In addition, the frequency of the most often mentioned activities participants reported feeling restricted in could be used to formulate a list of proposed activities for a condition specific PSFS 2.0. Additionally it can assist clinicians in their history taking and clinical reasoning process of patients with a CR.

With regards to the layout, it was of interest to see that none of the participants had comments concerning the layout of either questionnaire. Nevertheless, while completing the CRIS subscale 3, five participants initially missed completing the third question and only realized this while considering their rating on the fourth question.

The think aloud process has provided interesting insight into some of the participant's thoughts while reading and completing both the PSFS 2.0 as well as the CRIS subscale 3. Apart from some participants not identifying the example 11-point NRS scale in the PSFS 2.0 as an example and the above-mentioned initial accidental skipping of the second question on the CRIS subscale 3, some items apparently had ambiguous formulations (as mentioned in some of the quotes above) which might be handled easily in revised versions. This highlights the added value of qualitative and think aloud methods in developing PROMs. A study on the added value of qualitative studies in the development of health-related PROMs reported that participants with subacute neck pain distinguished six different types of problems: long complicated formulations, composite questions, irrelevant questions, lacking frame of reference, problematic words, and wrongly interpreted questions. It also suggested that qualitative methods have an added value when developing self-report questionnaires because some of the problems that were highlighted cannot be identified using quantitative methods only. They also recommend that a full qualitative study should be an integral part of the development of questionnaires (Pool et al., 2010). This is also relevant for item banks used in the PROMIS® (Turner-Bowker et al., 2012).

4.1. Strengths and limitations

One of the strengths of this study was that it adhered to the COREQ criteria (Tong et al., 2007). The structured use of a strict think aloud protocol while allowing for the evolution occurring during the process was an additional strength. We also feel that the combining of collecting and analysing both qualitative and quantitative data is innovative and adds important detail.

Having the participants read the entire PROMs out loud while completing them, including the instructions on how to complete them, provided little insight into the comparative length of time needed to complete them. Making them consciously consider all twenty proposed activities from the PSFS 2.0 skewed that in favour of the CRIS subscale 3.

Even though the NDI might not be the best PROM to assess functional limitations in patients with CR, in hindsight it might have been interesting to assess correlations between all three PROMs.

5. Conclusion

Both the PSFS 2.0 and the CRIS subscale 3 capture participation restrictions and functional limitations in patients with CR; both are easy to complete and take little time. The majority of patients included in this study prefer the PSFS 2.0 over the CRIS subscale 3 as a PROM. Some items of both PROMs have ambiguous formulations which might be handled easily in revised versions. Additionally, the layout of both PROMs need refinement to enhance user-friendliness and avoid misinterpretation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.msksp.2023.102764.

References

Ailliet, L., Knol, D.L., Rubinstein, S.M., de Vet, H.C., van Tulder, M.W., Terwee, C.B., 2013. 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. 66 (7), 775–782 quiz 782.e771-772.

Ailliet, L., Rubinstein, S.M., de Vet, H.C., van Tulder, M.W., Terwee, C.B., 2015. Reliability, responsiveness and interpretability of the neck disability index-Dutch version in primary care. Eur. Spine J. 24 (1), 88–93.

Alonso, J., Bartlett, S.J., Rose, M., Aaronson, N.K., Chaplin, J.E., Efficace, F., Leplège, A., Lu, A., Tulsky, D.S., Raat, H., Ravens-Sieberer, U., Revicki, D., Terwee, C.B., Valderas, J.M., Cella, D., Forrest, C.B., 2013. The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. Health Qual. Life Outcome 11, 210.

Ayub, A., Osama, M., Ahmad, S., 2019. Effects of active versus passive upper extremity neural mobilization combined with mechanical traction and joint mobilization in females with cervical radiculopathy: a randomized controlled trial. J. Back Musculoskelet. Rehabil. 32 (5), 725–730.

Chan, Y.H., 2003. Biostatistics 104: correlational analysis. Singap. Med. J. 44 (12), 614–619.

Cleland, J.A., Fritz, J.M., Whitman, J.M., Palmer, J.A., 2006. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine 31 (5), 598–602.

Coenen, M., Stamm, T.A., Stucki, G., Cieza, A., 2012. Individual interviews and focus groups in patients with rheumatoid arthritis: a comparison of two qualitative methods. Qual. Life Res. 21 (2), 359–370.

Ernstsson, O., Burström, K., Heintz, E., Mølsted Alvesson, H., 2020. Reporting and valuing one's own health: a think aloud study using EQ-5D-5L, EQ VAS and a time trade-off question among patients with a chronic condition. Health Qual. Life Outcome 18 (1), 388.

Fonteyn, M.E., Kuipers, B., Grobe, S.J., 1993. A description of think aloud method and protocol analysis. Qual. Health Res. 3 (4), 430–441.

Gale, N.K., Heath, G., Cameron, E., Rashid, S., Redwood, S., 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med. Res. Methodol. 13 (1), 117.

Gardner, B., Tang, V., 2014. Reflecting on non-reflective action: an exploratory think-aloud study of self-report habit measures. Br. J. Health Psychol. 19 (2), 258–273.

Gartner, F.R., Marinus, J., van den Hout, W.B., Vleggeert-Lankamp, C., Stiggelbout, A.M., 2019. The Cervical Radiculopathy Impact Scale: development and evaluation of a new functional outcome measure for cervical radicular syndrome. Disabil. Rehabil. 1–12.

Goransson, K., Ehrenberg, A., Ehnfors, M., Fonteyn, M., 2006. The use of qualitative data analysis software (QDAS) to manage and support the analysis of think aloud (TA) data. Stud Health Technol Inform 122, 143–146.

Guest, G., Bunce, A., Johnson, L., 2006. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 18 (1), 59–82.

Hassan, F., Osama, M., Ghafoor, A., Yaqoob, M.F., 2020. Effects of oscillatory mobilization as compared to sustained stretch mobilization in the management of cervical radiculopathy: a randomized controlled trial. J. Back Musculoskelet. Rehabil. 33 (1), 153–158.

Henry, G., Butow, P., Juraskova, I., Laidsaar-Powell, R., Shaw, J., 2022. Measuring female breast cancer survivors' concerns about late effects of treatment: initial development of the Concerns about Late Effects in Oncology Questionnaire (CLEO). Support. Care Cancer 30 (4), 3563–3572.

Horn, K.K., Jennings, S., Richardson, G., Vliet, D.V., Hefford, C., Abbott, J.H., 2012. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. J. Orthop. Sports Phys. Ther. 42 (1), 30–42.

Hudak, P.L., Amadio, P.C., Bombardier, C., 1996. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am. J. Ind. Med. 29 (6), 602–608.

Jette, D.U., Halbert, J., Iverson, C., Miceli, E., Shah, P., 2009. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys. Ther. 89 (2), 125–135.

Jordan, A., Begen, F.M., Austin, L., Edwards, R.T., Connell, H., 2020. A usability and feasibility study of a computerized version of the Bath Adolescent Pain Questionnaire: the BAPQ-C. BMC Pediatr. 20 (1), 6.

Kabboord, A.D., van Eijk, M., van Dingenen, L., Wouters, M., Koet, M., van Balen, R., Achterberg, W.P., 2019. Reliability and usability of a weighted version of the functional comorbidity index. Clin. Interv. Aging 14, 289–299.

Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., Bhattacharya, A., Gimson, A., Brütt, A.L., de Jong, K., Moore, M., 2016. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst. Rev. 7 (7), Cd011119.

Kuijper, B., Tans, J.T., Schimsheimer, R.J., van der Kallen, B.F., Beelen, A., Nollet, F., de Visser, M., 2009. Degenerative cervical radiculopathy: diagnosis and conservative treatment. A review. Eur. J. Neurol. 16 (1), 15–20.

Kuijper, B., Tans, J.T.J., van der Kallen, B.F., Nollet, F., Nijeholt, G., de Visser, M., 2011. Root compression on MRI compared with clinical findings in patients with recent onset cervical radiculopathy. J. Neurol. Neurosurg. Psychiatry 82 (5), 561–563.

Liang, L., Feng, M., Cui, X., Zhou, S., Yin, X., Wang, X., Yang, M., Liu, C., Xie, R., Zhu, L., Yu, J., Wei, X., 2019. The effect of exercise on cervical radiculopathy: a systematic review and meta-analysis. Medicine (Baltim.) 98 (45), e17733.

Lundgren-Laine, H., Salantera, S., 2010. Think-aloud technique and protocol analysis in clinical decision-making research. Qual Health Res 20 (4), 565–575.

M Granger Morgan, B.F., Bostrom, A., Atman, C., 2005. Risk Communication: A Mental Models Approach. Cambridge University Press.

Mehta, S., Macdermid, J.C., Carlesso, L.C., McPhee, C., 2010. Concurrent validation of the DASH and the QuickDASH in comparison to neck-specific scales in patients with neck pain. Spine 35 (24), 2150–2156.

Mjåset, C., Zwart, J.A., Goedmakers, C.M.W., Smith, T.R., Solberg, T.K., Grotle, M., 2020. Criteria for success after surgery for cervical radiculopathy-estimates for a substantial amount of improvement in core outcome measures. Spine J. 20 (9), 1413–1421.

Moses, M.J., Tishelman, J.C., Stekas, N., Jevotovsky, D.S., Vasquez-Montes, D., Karia, R., Errico, T., Buckland, A., Protopsaltis, T.S., 2019. Comparison of patient reported outcome measurement information system with neck disability index and visual analog scale in patients with neck pain. Spine 44 (3), E162–e167.

Namey, E., Guest, G., McKenna, K., Chen, M., 2016. Evaluating bang for the buck: a cost-effectiveness comparison between individual interviews and focus groups based on thematic saturation levels. Am. J. Eval. 37 (3), 425–440.

Nehlin, C., Wennberg, M., Oster, C., 2018. How do people with mood and anxiety disorders perceive and interpret the Drinking Motives Questionnaire? A think-aloud study in a clinical setting. Addiction Sci. Clin. Pract. 13 (1), 7.

Nielsen, J., 1994. Estimating the number of subjects needed for a thinking aloud test. Int. J. Hum. Comput. Stud. 41 (3), 385–397.

Noble, H., Smith, J., 2015. Issues of validity and reliability in qualitative research. Evid. Base Nurs. 18 (2), 34–35.

Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K., 2015. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Pol Ment Health 42 (5), 533–544.

Patton, M.Q., 2014. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Sage publications.

Peolsson, A., Löfgren, H., Dedering, Å., Öberg, B., Zsigmond, P., Hedevik, H., Wibault, J., 2019a. Postoperative structured rehabilitation in patients undergoing surgery for cervical radiculopathy: a 2-year follow-up of a randomized controlled trial. J. Neurosurg. Spine 31 (1), 60–69.

Peolsson, A., Peterson, G., Hermansen, A., Ludvigsson, M.L., Dedering, Å., Löfgren, H., 2019b. Physiotherapy after anterior cervical spine surgery for cervical disc disease: study protocol of a prospective randomised study to compare internet-based neck-specific exercise with prescribed physical activity. BMJ Open 9 (2), e027387.

Pool, J.J., Hiralal, S.R., Ostelo, R.W., van der Veer, K., de Vet, H.C., 2010. Added value of qualitative studies in the development of health related patient reported outcomes such as the Pain Coping and Cognition List in patients with sub-acute neck pain. Man. Ther. 15 (1), 43–47.

Rauch, A., Cieza, A., Stucki, G., 2008. How to apply the international classification of functioning, disability and health (ICF) for rehabilitation management in clinical practice. Eur. J. Phys. Rehabil. Med. 44 (3), 329–342.

Ritchie, J., Lewis, J., Nicholls, C.M., Ormston, R., 2013. Qualitative Research Practice: A Guide for Social Science Students and Researchers. sage.

Roach, K.E., Budiman-Mak, E., Songsiridej, N., Lertratanakul, Y., 1991. Development of a shoulder pain and disability index. Arthritis Care Res. 4 (4), 143–149.

Rothrock, N.E., Kaiser, K.A., Cella, D., 2011. Developing a valid patient-reported outcome measure. Clin. Pharmacol. Ther. 90 (5), 737–742.

Sleijser-Koehorst, M.L.S., Coppieters, M.W., Epping, R., Rooker, S., Verhagen, A.P., Scholten-Peeters, G.G.M., 2020. Diagnostic Accuracy of Patient Interview Items and Clinical Tests for Cervical Radiculopathy (Physiotherapy).

Sleijser-Koehorst, M.L.S., Coppieters, M.W., Epping, R., Rooker, S., Verhagen, A.P., Scholten-Peeters, G.G.M., 2021. Diagnostic accuracy of patient interview items and clinical tests for cervical radiculopathy. Physiotherapy 111, 74–82.

Tashakkori, A., Teddlie, C., 2010. Sage Handbook of Mixed Methods in Social and Behavioral Research. SAGE publications.

Thoomes-de Graaf, M., Fernandez-De-Las-Penas, C., Cleland, J.A., 2019. The content and construct validity of the modified patient specific functional scale (PSFS 2.0) in individuals with neck pain. J. Man. Manip. Ther. 1–11.

Thoomes-de Graaf, M., Scholten-Peeters, W., Duijn, E., Karel, Y., de Vet, H.C., Koes, B., Verhagen, A., 2017. The responsiveness and interpretability of the shoulder pain and disability index. J. Orthop. Sports Phys. Ther. 47 (4), 278–286.

Thoomes, E., Ellis, R., Dilley, A., Falla, D., Thoomes-de Graaf, M., 2021. Excursion of the median nerve during a contra-lateral cervical lateral glide movement in people with and without cervical radiculopathy. Musculoskelet Sci Pract 52, 102349.

Thoomes, E.J., Scholten-Peeters, G.G., de Boer, A.J., Olsthoorn, R.A., Verkerk, K., Lin, C., Verhagen, A.P., 2012. Lack of uniform diagnostic criteria for cervical radiculopathy in conservative intervention studies: a systematic review. Eur. Spine J. 21 (8), 1459–1470.

Thoomes, E.J., van Geest, S., van der Windt, D.A., Falla, D., Verhagen, A.P., Koes, B.W., Thoomes-de Graaf, M., Kuijper, B., Scholten-Peeters, W.G.M., Vleggeert-Lankamp, C. L., 2018. Value of physical tests in diagnosing cervical radiculopathy: a systematic review. Spine J. 18 (1), 179–189.

Tong, A., Sainsbury, P., Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int. J. Qual. Health Care 19 (6), 349–357.

Turner-Bowker, D.M., Saris-Baglama, R.N., Derosa, M.A., Paulsen, C.A., 2012. Cognitive testing and readability of an item bank for measuring the impact of headache on health-related quality of life. Patient 5 (2), 89–99.

Veehof, M.M., Sleegers, E.J., van Veldhoven, N.H., Schuurman, A.H., van Meeteren, N.L., 2002. Psychometric qualities of the Dutch language version of the disabilities of the arm, shoulder, and hand questionnaire (DASH-DLV). J. Hand Ther. 15 (4), 347–354.

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

Vleggeert-Lankamp, C.L.A., Janssen, T.M.H., van Zwet, E., Goedmakers, C.M.W., Bosscher, L., Peul, W., Arts, M.P., 2019. The NECK trial: effectiveness of anterior cervical discectomy with or without interbody fusion and arthroplasty in the treatment of cervical disc herniation; a double-blinded randomized controlled trial. Spine J. 19 (6), 965–975.

Wainner, R.S., Fritz, J.M., Irrgang, J.J., Boninger, M.L., Delitto, A., Allison, S., 2003a. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine 28 (1), 52–62.

Wainner, R.S., Fritz, J.M., Irrgang, J.J., Boninger, M.L., Delitto, A., Allison, S., 2003b. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine 28 (1), 52–62.

Witter, J.P., 2016. The promise of patient-reported outcomes measurement information system-turning theory into reality: a uniform approach to patient-reported outcomes across rheumatic diseases. Rheum. Dis. Clin. N. Am. 42 (2), 377–394.

Wu, P.H., Kim, H.S., Lee, Y.J., Kim, D.H., Lee, J.H., Yang, K.H., Raorane, H.D., Jang, I.T., 2020. Posterior endoscopic cervical foramiotomy and discectomy: clinical and radiological computer tomography evaluation on the bony effect of decompression with 2 years follow-up. Eur. Spine J. 457.

Wu, T.K., Liu, H., Ding, C., Rong, X., He, J.B., Huang, K.K., Hong, Y., Wang, B.Y., 2020. Effect of preoperative segmental range of motion on patient outcomes in cervical disc arthroplasty. BMC Muscoskel. Disord. 21 (1), 457.

Xu, X., Wang, Y., Yang, C., Song, X., Chen, Z., Yang, L., Li, Y., 2020. Evaluation of rehabilitation effect of five-step exercises on patients with radiculopathy of cervical vertebra. Medicine (Baltim.) 99 (26), e20846.

Zacher, S., Berger-Höger, B., Lühnen, J., Steckelberg, A., 2022. Development and piloting of a web-based tool to teach relative and absolute risk reductions. Int. J. Environ. Res. Publ. Health 19 (23).

Zhang, Y., Ouyang, Z., Wang, W., 2020. Percutaneous endoscopic cervical foraminotomy as a new treatment for cervical radiculopathy: a systematic review and meta-analysis. Medicine (Baltim.) 99 (45), e22744.