UNIVERSITY OF BIRMINGHAM

Inter-rater reliability, discriminatory and predictive validity of neck movement control tests in office workers with headache and/or neck pain

NEXpro collaboration group

[Link to publication on Research at Birmingham portal](#)

Original article

Inter-rater reliability, discriminatory and predictive validity of neck movement control tests in office workers with headache and/or neck pain

Markus J. Ernst [a,b,*], Sandro Klaus [b], Kerstin Lüdtke [c], Alessio Gallina [a], Deborah Falla [a], on behalf ofthe NEXpro collaboration group

[a] *Centre of Precision Rehabilitation for Spinal Pain, School of Sport, Exercise & Rehabilitation Sciences, University of Birmingham, Birmingham, UK*
[b] *Institute of Physiotherapy, School of Health Sciences, Zurich University of Applied Sciences, Katharina Sulzer Platz 9, 8400, Winterthur, Switzerland*
[c] *Institute of Health Sciences, Department of Physiotherapy, University of Luebeck, Germany*

1. Background

Headache conditions are on the rise, especially among females and in countries with high socio-demographic indices (Stovner, 2018). In 2016, the worldwide age-standardized prevalence for tension-type headache was 26.1% and for migraine, 14.4% (females: 30.8% and 18.9% respectively) (Stovner, 2018). Headache is the most prevalent neurological condition in Europe (Deuschl et al., 2020), and between 1990 and 2019, the worldwide health burden due to headache conditions has increased to rank 15 among all diseases worldwide (Vos, 2020). Office workers have a high prevalence of headache (Andersen et al., 2011; Rota et al., 2016). Many factors such as poor posture (Mingels et al., 2016), long working hours with increased physical inactivity and sleep deprivation have been associated with the presence of some headache conditions in office workers (Nagaya et al., 2018). Remote working due to the pandemic might have had additional detrimental effects (Houle et al., 2021).

Neck pain (NP) is frequently associated with headache conditions (Ashina et al., 2014; Al-Khazali et al., 2022), although the role of the cervical spine as either the source, coexisting factor, or an area of referred pain is controversial (Antonaci et al., 2001; Liang et al., 2021). The anatomical explanation for a reciprocal influence refers to convergence of afferences from upper cervical structures with trigeminal

afferents within the trigeminocervical nucleus (Edvinsson et al., 2020). With peripherally or centrally sensitized structures in the cervical spine, *positive* findings during the physical examination that exclusively rely on pain or muscle responses, have limited validity, as the "SpPIn rule" = specificity rules in, cannot be applied with low *specificity test values* (Liang et al., 2019). While according to the "*SnNOut rule*" = Sensitivity rules out, a test with high *sensitivity* which is *negative* is especially useful to rule *out* a cervical contribution (Davidson 2002; Luedtke and May, 2017; Baeyens et al., 2019).

According to the "movement control framework", movement control tests (MCTs) are intended to evaluate a potential loss of movement control, that can impair movement in the long-term (Dingenen et al., 2018, Mottram et al., 2020). MCTs are used to test the "control" of a specific "site" and "movement direction", and whether impaired control is related to symptoms, disability or pathology (Comerford and Mottram, 2012; Mottram et al., 2020). Although MCTs are frequently used to assess people with NP and headache, measurement properties are still lacking, and current evidence is limited to either rater-reliability (Patroncini et al., 2014; Segarra et al., 2015), or case-control studies that established "known-group validation" (Elsig et al., 2014; Aasa et al., 2020).

There is limited evidence whether individuals with headache differ from asymptomatic controls based on sensorimotor control tests

([Hodges and Falla, 2015](#)), such as the cranio-cervical-flexion test (CCFT) or joint position error test ([Szikszay et al., 2019](#); [Anarte-Lazo et al., 2021](#)), whereas a recent study demonstrated that motor control and movement accuracy differed between migraine participants with and without musculoskeletal dysfunction ([Liang et al., 2021](#)).

Establishing measurement properties of MCTs for the assessment of people with NP or headache is relevant as these tests may help to find subgroups with headache or NP that may be amenable to tailored exercise programmes, as has already been shown for MCTs in the lumbar region ([Luomajoki et al., 2018](#)) and for motor control interventions in NP ([Falla et al., 2012](#)). Thus, the aim of the current study was to establish the inter-rater reliability in addition to discriminatory and predictive validity for seven MCTs of the upper (UCS) and lower cervical spine (LCS) in office workers with and without headache or NP.

2. Methods

This is a validity and inter-rater reliability study. Baseline and final (15-month) follow-up data of a cluster randomized controlled trial, that sought to examine the effectiveness of exercise and health promotion on work presenteeism or absenteeism in office workers, were used for the evaluation of validity in this study ([Aegerter, 2020](#)). In a separate cross-sectional sample, inter-rater reliability was assessed. The study was approved by the ethical committee of the Canton Zurich
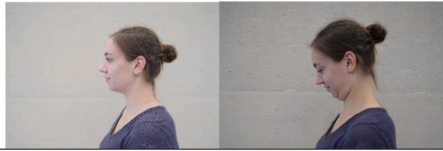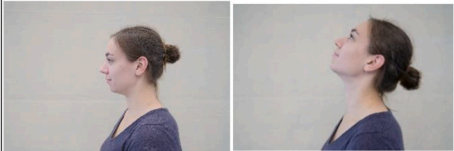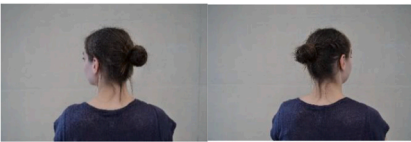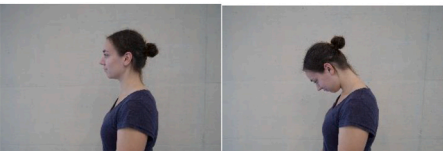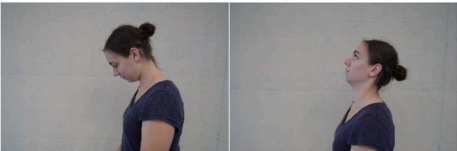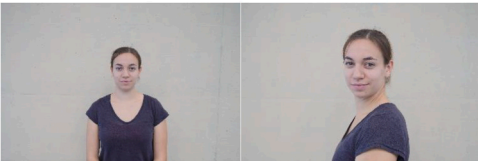
Movement control test	Instruction	Criteria for a «positive» rating
Upper cervical flexion control (UCS-F)	Bend forward only your head, not your neck, like in nodding.	Moves and initiates movement dominantly in the mid and lower cervical spine.
Upper cervical extension control (UCS-E)	Lift your chin up, do not push it forward.	Moves and initiates movement dominantly in the mid and lower cervical spine. Protrudes chin forward, leading to mid cervical flexion.
Upper cervical rotation (UCS-Rot)	Turn only your head, not your neck to left and right, like in shaking your head	Movement with concomitant and predominantly lateroflexion. Movement continues into mid and lower cervical spine. Movement below 20° by visual inspection
Lower cervical spine flexion (LCS-F)	Bend your neck forward, with your head stable, not nodding or protruding the head. Imagine the pivoting point is in the neck, not head. Think about your head being a Ball on a stick. Bend the stick, while not spinning the ball.	Movement with pivoting predominantly in the head. Protrudes chin forward and downwards during the test.
Lower cervical spine extension	Start in approx. 30° Flexion. Lift your head and neck towards an upright posture and slightly further. Initiate the movement in the lower cervical spine by rolling up from there	Initiates and moves in the upper cervical spine by moving the chin first.
Cervical spine rotation left and right	Turn your head to the side (left or right). Imagine you spin around an axis that points down your spine.	Movement with concomitant and predominantly lateroflexion or protrusion. Movement below 60° to one side (taking the age of the subject into account).

Fig. 1. Movement control tests 220514.

(Ref-No.2019-01678) and registered at clinicaltrials.gov (NCT04169646). All participants gave written informed consent before taking part. "Standards for the Reporting of Diagnostic Accuracy studies" (STARD) have been followed (https://www.equator-network.org/).

2.1. Participants

Adult office workers with at least 25 sedentary working hours per week were included. Exclusion criteria were self-reported specific health conditions: NP grade III and IV according to the European taskforce criteria (Haldeman et al., 2008), general inflammatory conditions such as rheumatoid arthritis, spinal surgery or other conditions that prevented participants from exercising. Being non-fluent in German, planned (>4 weeks) absences during the intervention period and pregnancy were further exclusion criteria.

2.2. Movement control tests

Three MCTs for the (UCS) and four MCTs for the (LCS) were examined at baseline of the trial (Fig. 1). Selected tests provide information on the control of movement in both regions, UCS and LCS, and for two directions: sagittal plane (four tests) and horizontal plane (three tests) (Comerford and Mottram, 2012; Segarra et al., 2015). For consistency in ratings, tests were performed in a pre-determined order: UCS before LCS, and sagittal (flexion = F before extension = E) before horizontal (rotation = Rot). Positive ratings (participant could not perform a test correctly) were added to obtain a total score (range from 0 to 7 across all seven tests), and sub-scores were obtained to differentiate UCS control (range 0–3) from LCS control (range 0–4) and sagittal plane control (range 0–4) from rotational control (range 0–3). First, participants received verbal instructions on how to perform each test (Fig. 1). Verbal feedback was provided, and standardized videos of each test had been shown (supplemental files), if participants did not understand the task. Participants had sufficient time and repetitions to familiarize themselves with the tests, until they considered themselves capable to perform the test correctly. Eventually, each test was performed without feedback and rated as either positive or negative (criteria provided in Fig. 1).

2.3. Reliability

MCTs were examined by a random pair of two, out of a pool of seven raters, that also participated in the validity study. Accordingly, only inter-rater reliability was examined, to assure consistent ratings for the subsequent validity study. Raters received training to be able to detect a "lack of control" for each test, according to criteria by Comerford & Mottram and described in Fig. 1 (Comerford and Mottram, 2012). Raters were kept blind to each other's ratings and to information about the headache and/or NP status of the participant. Information on raters can be found in Table 1.

2.4. Validity

Participants completed an online questionnaire, that screened and assessed for potential headache (IHS 2018) or NP in the last four weeks. That questionnaire included the Numeric pain rating scale (Jensen et al.,

Table 1
Information on raters.

Variable	Values
Sex (female/male)	4/3
Age in years	35.6 (9.2)
Graduated in years	11.4 (9.7)
Basic vs. specific training (IFOMPT certified)	5/2

Values are means (sd) and absolute frequencies (/).

1986), the Headache-impact-test −6 (HIT-6) (Haywood et al., 2018) and the Neck disability index (NDI) (Swanenburg et al., 2014). Baseline data on the occurrence of headache or NP was used to determine the *discriminatory validity* of MCT between different subgroups (De Vet et al., 2011). Due to the nature of the study design as a stepped wedge cluster RCT (Hemming et al., 2015), MCTs and questionnaires for all participants had to be completed within two to three weeks in January 2020, which necessitated the inclusion of seven raters. After the inter-rater reliability study, raters received additional training and videos of correct performance (supplemental files). MCTs were performed as described, and were rated by one out of seven raters, who was blind to the headache/NP status of the participant.

At three-monthly online follow-up assessments, the occurrence of headache or NP within the last 4 weeks were re-examined. For this study, data from the last follow-up after 15 months was used to examine the *predictive validity of MCT performance*, when all participants would have received the intervention.

2.5. Data analysis

Inter-rater-reliability data were analysed using Generalisability theory (G-theory). G-theory is based on classical test theory and analysis of variance (ANOVA). Variances of influencing factors (rater, subject) contribute to a G-coefficient, that is similar to an intra-class correlation coefficient (ICC) (Brennan 2011). As subjects are regarded "nested" in raters, results for varying pairs of raters contribute to the same coefficient. Since a low prevalence of positive findings is known to negatively affect coefficients, irrespective of rater agreement, "prevalence-adjusted coefficients" were additionally calculated (Sim et al., 2005). G-coefficients of <0.2 were interpreted as slight, 0.21–0.4 as fair, 0.41–0.6 as moderate, 0.61–0.8 as substantial, 0.81–1 as almost perfect (Landis and Koch, 1977).

For *discriminatory* and *predictive validity*, the area under the curve (AUC) was calculated to find optimal cut-off points for composite scores such as for the total MCT score, regional or directional scores (Baeyens et al., 2019). Only for AUC scores ≥0.5, positive (LR+) and negative (LR-) likelihood ratios, together with diagnostic odds ratios (DORs) were computed for *discriminatory*, and together with relative risk ratios (RRs) for *predictive* validity (Davidson 2002; Knottnerus et al., 2008). A LR expresses the likelihood of a positive or negative test result in someone with or without, respectively, the problem under investigation (Davidson 2002; Knottnerus et al., 2008). LRs are more meaningful than sensitivity or specificity values alone (Baeyens et al., 2019). A LR+ of >10 is regarded as "high discriminatory or predictive value", between 5 and 10 is of "moderate value", below 5 but at least 2 is of "small value", and below 2 but above 1 is of "limited value". For LR-, a "high value" would be regarded as < 0.1, "moderate value" as 0.1 to 0.2, a "small value" for values > 0.2–0.5, and a "limited value" for >0.5 to 1 (Sleijser-Koehorst et al., 2021). A DOR is the ratio between LR+ and LR- and can be interpreted as "the probability of having the problem in someone with a positive test compared to someone with a negative test". A relative risk ratio (RR) expresses the "likelihood that someone with a positive test compared to someone with a negative test result at baseline will have the problem at follow-up" (Gross Portney and Watkins, 2000). 95% confidence intervals (95% CI) for all ratios have been calculated; a value of 1 indicates "no effect". Computations have been performed for the entire cohort and for contrast between the following subgroups with: "headache only", "NP only", "headache & NP", and "asymptomatic" for both conditions. Furthermore, subgroups with "headache only" PLUS "headache & NP" were merged to a larger "all headache" subgroup, as were "NP only" PLUS "headache & NP" to "all NP". On the opposite those with "NP only" PLUS "asymptomatic" were merged to "no headache", as were those with "headache only" PLUS "asymptomatic" to "no NP". Missing values at the 15-months follow-up were replaced by the "last observation carried forward" method (Uthaikhup et al., 2017).

3. Results

In total 140 office workers participated in both parts of the study. Descriptive data of participants is presented in Table 2.

3.1. Inter-rater reliability

Prevalence adjusted G-coefficients for MCTs ranged from slight (UCS-Rot.) to almost perfect (UCS–F, LCS-F, LCS-E), (Table 3).

3.2. Validity

One hundred and twelve of 120 participants reported about headache and/or NP at baseline. Seventy-one participants of all those who were symptomatic suffered from both headache and NP at baseline. See Fig. 2.

The percentage of participants that scored positive on individual MCTs were: UCS-F (15%), UCS-E (29.2%), UCS-Rot 17.5%, LCS-F (44.2%), LCS-E (54.2%) and Rot left and right (each 27.5%).

Discriminatory validity: In total, six of eight possible contrasts between subgroups suffering from headache or NP had an AUC ≥0.5. Single or composite MCTs had limited validity to discriminate between subgroups at baseline. For the UCS ≤1/3 positive MCTs are more likely in "no headache" than in "all headache" (LR-: 0.89, 95% CI: 0.80–0.99). Furthermore, UCS-Rot control had limited validity to discriminate between "all NP" and "no NP" (LR-: 0.82, 95% CI: 0.72–0.94), and between "headache & NP" versus "headache only" (LR-: 0.82, 95% CI: 0.69–0.98). Further results can be found in Table 4.

Predictive validity: Single or composite MCTs had limited to small validity to predict a future headache or NP event. At the 15-month follow-up, a headache event was more likely in participants with baseline "headache & NP", and >2/7 positive MCTs of the total score (RR-: 2.47, 95% CI: 1.17–5.22).

A NP event at follow-up was less likely in participants with "NP only" at baseline AND 0/4 positive MCTs in the sagittal plane (LR-: 0.30, 95% CI: 0.09–0.95). The RR for a future NP event in this subgroup AND ≥1/4

Table 2
Sample descriptives.

Variable	Reliability study	Validity study		
		Baseline n = 120	**15 months follow-up**	
			All (ITT)	Completers (n = 91)
Participants (females)	20 (14)	120 (86)		
Age	46.2 (9.9)	44.2 (9.8)		
Headache in the last 4 weeks ≥1 occurrence (%)	13 (65)	88 (73.3)	71 (59.2)	55 (60.4)
Neck pain in the last 4 weeks ≥1 occurrence (%)	NA	95 (79.2)	70 (58.3)	58 (63.7)
Number of headache days in the last 4 weeks (median/iqr)	2.0 (1.88–3.25)	2 (0–5)	1.75 (0–5)	1.5 (1–4)
Number of neck pain days in the last 4 weeks (median/iqr)	NA	4 (1–8)	2 (0–7)	2.25 (0–8)
Average headache intensity (0–10)	4.8 (2.9)	4.3 (2.2)	NA	NA
Average neck pain intensity (0–10)	NA	3.0 (1.8)	2.5 (1.4)	2.8 (1.5)
HIT-6 score (36–78 points)	NA	53-6 (7.9)	51.8 (6.7)	51.9 (6.2)
NDI score (0–50 points)	5.7 (3.9)	5.8 (5.0)	7.2 (4.6)	8.1 (4.7)

Values are mean (sd) otherwise indicated; HIT-6 (Headache impact test); NDI= Neck disability index; iqr = interquartile range, ITT = intention to treat analysis; NA= Not available.

positive MCTs in the sagittal plane was 3.33 (95% CI 1.05–10.56). Further results can be found in Table 5.

4. Discussion

This study investigated the *inter-rater reliability*, *discriminatory* and *predictive* validity of seven MCTs of the cervical spine in a cohort of office workers with and without headache and/or NP. We found slight to almost perfect inter-rater reliability, limited discriminatory validity, and limited to small predictive validity for some tests to identify office workers with a current or future headache or NP condition.

While six MCTs achieved, after prevalence-adjustment, a moderate to almost perfect inter-rater reliability, UCS-Rot control showed very low inter-rater reliability, which may also have compromised its validity (Table 3). Reasons for this low inter-rater reliability might be that simultaneous movements in the LCS were difficult to judge, as the rotation movement axis is difficult to visualise, compared to sagittal plane movements, with an imaginary axis approximately through the ears. Skin creases might help to visualise a pivoting movement are usually less obvious during rotation. A previous study, reporting on UCS-Rot control, performed in a 4-point kneeling position, found substantial rater reliability when testing people with NP (Segarra et al., 2015). Elsig et al. also demonstrated the validity of rotation control in a 4-point kneeling position to discriminate between people with and without NP, but did not limit the movement to the UCS (Elsig et al., 2014). In both studies, rotation was performed against gravity, possibly explaining better reliability as higher loading conditions can accentuate a movement control impairment, while our study showed only 17.5% scored positive for UCS-Rot (Mottram et al., 2020). Most of the previous studies on movement control testing at the cervical spine have either examined intra-or inter-rater reliability (Aasa et al., 2014; Patroncini et al., 2014; Segarra et al., 2015; Tegern et al., 2018) or validity to discriminate "known groups" of NP participants from asymptomatic controls (Elsig et al., 2014; Aasa et al., 2020).

No study so far had examined its *discriminatory validity* in symptomatic subjects with differing complaints, such as NP and/or headache. Elsig et al. reported exclusively on NP participants with symptoms indicating towards a movement control deficit, but compared them with asymptomatic controls, which may overestimate the value of MCTs for clinical settings (Elsig et al., 2014).

No study had examined the *validity* of MCTs to *predict* future headache or NP events. Furthermore, previous studies provided little information about a specific region or direction of impaired movement control, that might help direct patient management (Sahrmann 2010, Comerford and Mottram, 2012). Khosrokinani et al. used movement control as an intervention in computer workers with chronic NP and found large effects on NP-related disability (Khosrokiani et al., 2018). It remains unknown whether participants had been affected by a movement control impairment. In people suffering from low back pain, subgrouping according to their movement control impairment showed limited long-term effects for pain and disability (Luomajoki et al., 2018).

Movement control requires proprioception, central processing, including matching of stored motor pattern with task requirements, and an adequate motor output (Kristjansson and Treleaven, 2009; Wallwork et al., 2020). Furthermore, individual, environmental, and task-specific factors can influence someone's movement strategy (Mottram et al., 2020). In the current study, task-specific and environmental factors were standardized. Furthermore, like Segarra et al., MCTs were performed after participants indicated they had understood the movement task, which means, not testing primarily, a lack of understanding (Segarra et al., 2015; Dingenen et al., 2018). All tests were performed in a sitting position leading to gravity induced loading conditions for sagittal plane movements (Dingenen et al., 2018, Mottram et al., 2020).

From a clinical perspective the most important findings are that participants with "headache & NP" differed from "asymptomatic", and from "headache only", by a more frequently positively rated UCS-Rot

Table 3
Interrater reliability (n = 20).

Test	UCS-F control	UCS-E control	UCS-Rot control	LCS-F control	LCS-E control	Rot-L control	Rot-R control
G-coefficient (95% CI)	0.99 (0.9–1)	0.39 (0–1)	0.00 (0–0.43)	0.99 (0.98–1)	0.85 (0.43–1)	0.24 (0–0.88)	0.95 (0.72–1)
Prevalence adjusted G-coefficient (95% CI)	0.93 (0.67–1)	0.57 (0.16–1)	0.16 (0–0.54)	0.99 (0.98–1)	0.85 (0.41–1)	0.43 (0.02–1)	0.70 (0.29–1)
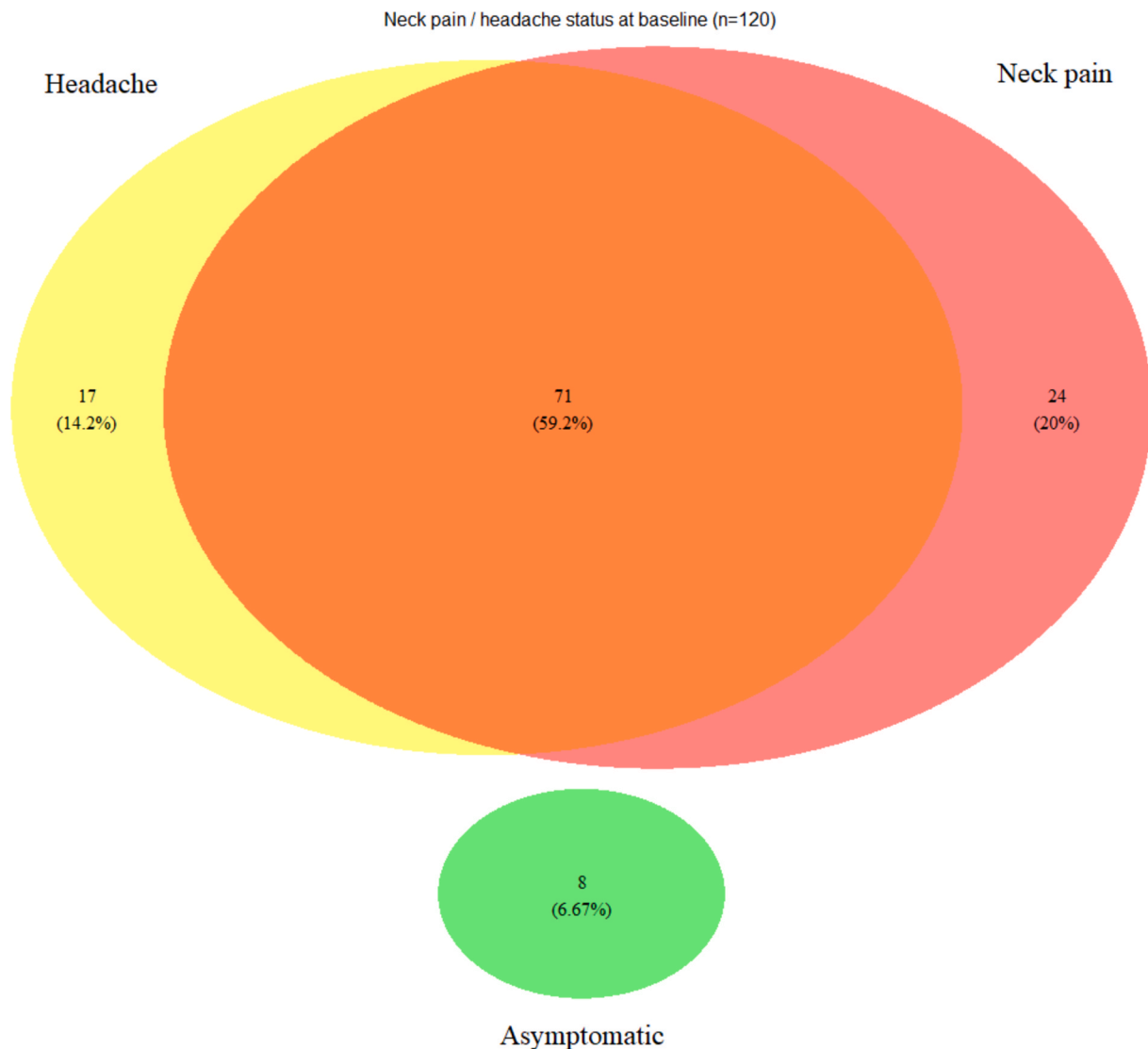
Fig. 2. Venn diagram of Neck pain/headache status at baseline.

Table 4
Discriminatory validity.

Contrast	Test (cut-off)	AUC	LR+	LR-	DOR
"All headache" (n = 88) VS. "no headache" (n = 32)	UCS control (>1/3)	0.51	4.36 (0.59–32.23)	**0.89 (0.80–0.99)**	4.85 (0.66–215.68)
"All NP" (n = 95) VS. "no NP" (n = 25)	Rotation control (>0/3)	0.58	1.14 (0.95–1.38)	0.60 (0.29–1.23)	1.88 (0.70–5.21)
	UCS-Rot control	0.60	5.26 (0.74–37.34)	**0.82 (0.72–0.94)**	6.34 (0.91–275.82)
"Headache & NP" (n = 71) VS "headache only" (n = 17)	UCS-Rot control	0.58	3.83 (0.55–26.92)	**0.82 (0.69–0.98)**	4.60 (0.62–207.02)
"Headache & NP" (n = 71) VS. "NP only" (n = 24)	UCS control (>1/3)	0.51	2.70 (0.36–20.52)	0.93 (0.82–1.04)	2.90 (0.35–135.07)
"Headache & NP" (n = 71) VS. "asymptomatic" (n = 8)	UCS-Rot control	0.61	2.33 (0.35, 15.65)	0.85 (0.67, 1.09)	2.71 (0.33–126.63)
"NP only" (n = 24) VS "headache only" (n = 17)	UCS-Rot control	0.55	2.83 (0.35, 23.17)	0.89 (0.71, 1.10)	3.12 (0.27–167.12)

AUC = area under the curve: (only contrasts with ≥0.5 are shown); LR+: positive likelihood ratio; LR-: negative likelihood ratio; DOR: diagnostic odds ratio; LCS: lower cervical spine; NP = neck pain; Rot = rotation; UCS: upper cervical spine; Values in brackets are 95% confidence intervals. Values in **bold** do not include the effect of the null hypothesis.

Table 5
Predictive validity for outcomes at 15-months follow-up.

Prediction	(Sub-)sample	Test (cut-off)	AUC	LR+	LR-	RR
HEADACHE in last 4 weeks,	Entire cohort (n = 120)	None	NA	NA	NA	NA
	Subgroup: "all headache" at baseline (n = 88)	Total control (>1/7)	0.56	1.37 (0.94–2.00)	0.57 (0.32–1.02)	1.76 (0.97–3.20)
		LCS control (>0/4)	0.55	1.16 (0.91–1.46)	0.53 (0.21–1.32)	1.62 (0.85–3.11)
		Sagittal control (>0/4)	0.57	1.22 (0.90–1.66)	0.61 (0.30–1.21)	1.57 (0.85–2.89)
		UCS-Rot control	0.59	2.18 (0.68–6.97)	0.86 (0.71–1.04)	2.00 (0.68–5.84)
	Subgroup: "all NP" at baseline (n = 95)	Total control (>1/7)	0.58	1.34 (0.98–1.85)	**0.54 (0.30–0.99)**	**1.71 (1.03–2.84)**
		LCS control (>0/4)	0.58	1.21 (0.96–1.53)	0.47 (0.20–1.07)	**1.71 (1.01–2.89)**
		Sagittal control (>0/4)	0.58	1.60 (0.80–3.21)	0.82 (0.63–1.07)	1.56 (0.81–3.01)
		LCS-F control	0.58	1.54 (0.88–2.68)	0.75 (0.54–1.05)	1.59 (0.88–2.85)
	Subgroup "headache & NP" at baseline (n = 71)	Total control (>2/7)	0.62	1.62 (0.99–2.67)	**0.44 (0.23–0.84)**	**2.47 (1.17–5.22)**
		LCS control (>0/4)	0.61	1.26 (0.92–1.75)	0.43 (0.16–1.11)	2.06 (0.97–4.40)
		Sagittal control (>1/4)	0.61	2.19 (0.73–6.61)	0.78 (0.59–1.02)	2.24 (0.73–6.89)
		LCS-F control	0.59	1.90 (0.85–4.23)	**0.68 (0.46–0.99)**	2.17 (0.88–5.38)
	Subgroup: "headache only" at baseline (n = 17)	UCS-Rot control	0.78	2.20 (0.23–20.72)	0.88 (0.61–1.26)	1.67 (0.32–8.70)
NECK PAIN in the last 4 weeks	Entire cohort (n = 120)	UCS-Rot control	0.55	0.93 (0.79–1.09)	1.30 (0.68–2.48)	1.30 (0.68–2.48)
	Subgroup: "all headache" at baseline (n = 88)	UCS-Rot control	0.61	2.25 (0.80–6.35)	0.84 (0.69–1.03)	1.92 (0.78–4.69)
	Subgroup: "headache only" at baseline (n = 17)	UCS-Rot control	0.84	3.25 (0.35–30.32)	0.81 (0.53–1.25)	2.0 (0.39–10.24)
	Subgroup: "NP only" at baseline (n = 24)	Total score (>2/7)	0.70	3.00 (0.43–20.86)	0.71 (0.45–1.13)	2.88 (0.43–19.30)
		LCS control (>1/4)	0.75	2.75 (0.79–9.55)	**0.42 (0.18–0.96)**	3.55 (0.89–14.15)
		Sagittal control (>0/4)	0.77	2.17 (0.86–5.46)	**0.30 (0.09–0.95)**	**3.33 (1.05–10.56)**
		LCS-E control	0.70	2.75 (0.79–9.55)	0.42 (0.18–0.96)	3.55 (0.89–14.15)
		LCS-F control	0.68	4.00 (0.60–26.68)	**0.57 (0.33–1.00)**	4.20 (0.61–28.80)

AUC = area under the curve: (only contrasts with ≥0.5 are shown); E = Extension; F= Flexion, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; RR = relative risk ratio; LCS = lower cervical spine; NP = neck pain; Rot = rotation; UCS = upper cervical spine; Values in brackets are 95% confidence intervals. Values in **bold** do not include the effect of the null hypothesis.

MCT. In addition, participants with "headache & NP" AND >2 of 7 positive MCTs were roughly 2.5 more likely to report a headache event in the future. These findings suggest that impaired cervical movement control may be a clinical feature for someone with both "headache & NP", especially if UCS-Rot control is among positively rated tests (Tables 4 and 5). Reduced UCS range of movement, and especially rotation has previously been reported in people suffering from headache related to the neck (Hall et al., 2010; Ernst et al., 2015; Bragatto et al., 2019). Restrictions in range of motion were not examined in the current study but might be associated to more frequent "positive" ratings during UCS-Rot in headache and NP.

Participants with "NP only", and poor movement control in the sagittal plane or lower CS (>1/4) had a 3-fold increased risk for future NP. (Table 5). Negative LRs showed more precise confidence intervals for most contrasts, accordingly it appears that MCTs, when tested negative may be better at *excluding* the problem under investigation. Similar findings have been reported by Rodrigues et al. for the cranio-cervical-flexion-test (CCFT), a test of deep neck flexor activity, to detect migraine subjects without disability (Rodrigues et al., 2021), and for pain provocation during manual palpation of upper cervical segments to exclude their involvement in the pain experience (Luedtke and May, 2017).

There are some limitations of this study. As participants had simultaneously entered and attended assessment of the trial, participation of seven raters was necessary. Inter-rater reliability for UCS-Rot was limited at best and may have compromised the results of the validity study. However, as preliminary results of the inter-rater-reliability study were known four weeks before beginning of the validity study (Aegerter, 2020), further training of raters was undertaken. Intra-rater reliability

was not evaluated. Since inter-rater reliability also incorporates errors of intra-rater reliability plus differences between raters, its evaluation is not considered a major limitation (Streiner and Norman, 2008).

All tests were performed in a standardized, pre-determined order. This approach has been used in previous studies (Aasa et al., 2014; Segarra et al., 2015) and is not regarded a major limitation of the current study. MCTs for most participants during the validity study were observed by two raters, and consent between raters had been sought, obtained values are regarded plausible and valid. Only office workers were examined, of whom many suffered from only mild NP (Table 2) or headache (Rendas-Baum et al., 2014). This likely limits the generalisability of the results towards the general population. All participants of the validity studies took part in a stepped wedge cluster RCT receiving health promotion and neck muscle exercises that focussed on strength and endurance of neck and shoulder girdle muscles (Aegerter, 2020). These interventions might have influenced movement control too, even if this was not the primary intention and should have been equal for all participants at the final follow-up. Further, sample sizes for some specific subgroups were rather small, hampering further adjustments for potential confounders, such as the headache type, and possibly leading to results with low precision (Elkins et al., 2022). On the other hand, insignificant results for predictive validity, even in larger subgroups such as "all headache types" might have been "washed out" by combining those suffering from "headache only" with those suffering also from NP (Liang et al., 2021). Regarding NP as the most prevalent associated symptom for many headache types, statistical adjustment was disregarded (Liang et al., 2019; Al-Khazali et al., 2022). Furthermore, movement variability has not only been found in healthy states but also in pain conditions and may have led to insignificant or imprecise results

for some contrasts (Dingenen et al., 2018). Last, each MCT contributed equally to a composite score, assuming that each test is equally difficult to control. However, positive findings were more often found for LCS MCTs and in the sagittal plane, indicating that these tests were generally more difficult to perform. Adelt et al. determined varying difficulties for directional MCTs for the lower back using an item-response theory approach (Adelt et al., 2021). They proposed an ordered use of tests for the same direction, starting with the most difficult (Adelt et al., 2021). In this respect, rotational control tests in the current study might have been too easy and may have led to negative scores, even in symptomatic participants.

In conclusion, movement control tests of the cervical spine have shown slight to nearly perfect inter-rater reliability. Impaired movement control, especially of UCS-Rot was found slightly more often in people with headache and NP. Risk for future headache events was higher in those with initial headache & NP AND >2/7 positive MCTs, while future NP events are more likely in those with initial NP and impaired movement control in the sagittal plane.

Declaration of interest

None.

Ethical approval

The study was approved by the Ethical committee of the Canton Zurich (Ref-No. 2019–01678) and was registered at clinicaltrials.gov (NCT04169646).

Funding

Acknowledgments:

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2022.102685.

References

Aasa, B., et al., 2014. Do we see the same movement impairments? The inter-rater reliability of movement tests for experienced and novice physiotherapists. Eur. J. Physiother. 16 (3), 173–182. https://doi.org/10.3109/21679169.2014.917435.

Aasa, B., et al., 2020. Acuity of goal-directed arm movements and movement control; evaluation of differences between patients with persistent neck/shoulder pain and healthy controls. Eur. J. Physiother. 24 (1), 47–55. https://doi.org/10.1080/21679169.2020.1785004.

Adelt, E., et al., 2021. Lumbar movement control in non-specific chronic low back pain: evaluation of a direction-specific battery of tests using item response theory. Musculoskelet. Sci. Pract. 55, 102406 https://doi.org/10.1016/j.msksp.2021.102406.

Aegerter, A.M., et al., 2020. On-site multi-component intervention to improve productivity and reduce the economic and personal burden of neck pain in Swiss office-workers (NEXpro): protocol for a cluster-randomized controlled trial. BMC Musculoskeletal Disorders 21 (1), 391. https://doi.org/10.1186/s12891-020-03388-x.

Al-Khazali, H.M., et al., 2022. Prevalence of neck pain in migraine: a systematic review and meta-analysis. Cephalalgia. https://doi.org/10.1177/03331024211068073, 3331024211068073.

Anarte-Lazo, E., et al., 2021. Differentiating migraine, cervicogenic headache and asymptomatic individuals based on physical examination findings: a systematic review and meta-analysis. BMC Muscoskel. Disord. 22 (1), 755. https://doi.org/10.1186/s12891-021-04595-w.

Andersen, L.L., et al., 2011. Effect of brief daily exercise on headache among adults – secondary analysis of a randomized controlled trial. Scand. J. Work. Environ. Health (6), 547–550. https://doi.org/10.5271/sjweh.3170.

Antonaci, F., et al., 2001. Cervicogenic headache: evaluation of the original diagnostic criteria. Cephalalgia 21 (5), 573–583. https://doi.org/10.1046/j.0333-1024.2001.00207.x.

Ashina, S., et al., 2014. Prevalence of neck pain in migraine and tension-type headache: a population study. Cephalalgia 35 (3), 211–219. https://doi.org/10.1177/0333102414535110.

Baeyens, J.-P., et al., 2019. Questioning the "SPIN and SNOUT" rule in clinical testing. Arch. Physiother. 9 (1), 4. https://doi.org/10.1186/s40945-019-0056-5.

Bragatto, M.M., et al., 2019. Is the presence of neck pain associated with more severe clinical presentation in patients with migraine? A cross-sectional study. Cephalalgia 39 (12), 1500–1508. https://doi.org/10.1177/0333102419854061.

Brennan, R.L., 2011. Generalizability theory and classical test theory. Appl. Meas. Educ. 24 (1), 1–21. https://doi.org/10.1080/08957347.2011.532417.

Comerford, M., Mottram, S., 2012. Kinetic Control. Churchill Livingstone, Chatswood, Australia.

Davidson, M., 2002. The interpretation of diagnostic tests: a primer for physiotherapists. Aust. J. Physiother. 48 (3), 227–232. https://doi.org/10.1016/S0004-9514(14)60228-2.

De Vet, H.C., et al., 2011. Measurement in Medicine: a Practical Guide. Cambridge university press.

Deuschl, G., et al., 2020. The burden of neurological diseases in Europe: an analysis for the global burden of disease study 2017. Lancet Public Health 5 (10), e551–e567. https://doi.org/10.1016/s2468-2667(20)30190-0.

Dingenen, B., et al., 2018. The assessment of movement health in clinical practice: a multidimensional perspective. Phys. Ther. Sport 32, 282–292. https://doi.org/10.1016/j.ptsp.2018.04.008.

Edvinsson, J.C.A., et al., 2020. The fifth cranial nerve in headaches. J. Headache Pain 21 (1), 65. https://doi.org/10.1186/s10194-020-01134-1.

Elkins, M.R., et al., 2022. Statistical inference through estimation: recommendations from the international society of physiotherapy journal editors. J. Physiother. 68 (1), 1–4. https://doi.org/10.1016/j.jphys.2021.12.001.

Elsig, S., et al., 2014. Sensorimotor tests, such as movement control and laterality judgment accuracy, in persons with recurrent neck pain and controls. A case-control study. Man. Ther. 19 (6), 555–561. https://doi.org/10.1016/j.math.2014.05.014.

Ernst, M.J., et al., 2015. Extension and flexion in the upper cervical spine in neck pain patients. Man. Ther. 20 (4), 547–552. https://doi.org/10.1016/j.math.2014.12.005.

Falla, D., et al., 2012. 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 28 (7), 628–634. https://doi.org/10.1097/AJP.0b013e31823e9378.

Gross Portney, L., Watkins, M.P., 2000. Foundations of Clincal Research, Application to Practice. Prentice Hall, New Jersey.

Haldeman, S., et al., 2008. The bone and joint decade 2000-2010 task force on neck pain and its associated disorders: executive summary. Spine (Phila Pa 1976) 33 (4 Suppl. l), S5–S7. https://doi.org/10.1097/BRS.0b013e3181643f40.

Hall, T.M., et al., 2010. Comparative analysis and diagnostic accuracy of the cervical flexion-rotation test. J. Headache Pain 11 (5), 391–397. https://doi.org/10.1007/s10194-010-0222-3.

Haywood, K.L., et al., 2018. Assessing the impact of headaches and the outcomes of treatment: a systematic review of patient-reported outcome measures (PROMs). Cephalalgia 38 (7), 1374–1386. https://doi.org/10.1177/0333102417731348.

Hemming, K., et al., 2015. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 350, h391. https://doi.org/10.1136/bmj.h391.

Hodges, P., Falla, D., 2015. Interaction between pain and sensorimotor control. In: Jull, G., Moore, A., Falla, D., et al. (Eds.), Grieve's Modern Musculoskeletal Physiotherapy. Elsevier, Edinburgh, pp. 53–57.

Houle, M., et al., 2021. Factors associated with headache and neck pain among telecommuters – a five days follow-up. BMC Publ. Health 21 (1), 1086. https://doi.org/10.1186/s12889-021-11144-6.

IHS, 2018. Headache classification committee of the international headache society (IHS) the international classification of headache disorders, 3rd edition. Cephalalgia 38 (1), 1–211. https://doi.org/10.1177/0333102417738202.

Jensen, M.P., et al., 1986. The measurement of clinical pain intensity: a comparison of six methods. Pain 27 (1), 117–126. https://doi.org/10.1016/0304-3959(86)90228-9.

Khosrokiani, Z., et al., 2018. Long-term effect of direction-movement control training on female patients with chronic neck pain. J. Bodyw. Mov. Ther. 22 (1), 217–224. https://doi.org/10.1016/j.jbmt.2017.06.004.

Knottnerus, J.A., et al., 2008. General introduction: evaluation of diagnostic procedures. Evid. Base Clin. Diag. 1–19.

Kristjansson, E., Treleaven, J., 2009. Sensorimotor function and dizziness in neck pain: implications for assessment and management. J. Orthop. Sports Phys. Ther. 39 (5), 364–377. https://doi.org/10.2519/jospt.2009.2834.

Landis, J.T., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Liang, Z., et al., 2019. Cervical musculoskeletal impairments in migraine and tension type headache: a systematic review and meta-analysis. Musculoskelet. Sci. Pract. 42, 67–83. https://doi.org/10.1016/j.msksp.2019.04.007.

Liang, Z., et al., 2021. Neck pain associated with migraine does not necessarily reflect cervical musculoskeletal dysfunction. Headache 61 (6), 882–894. https://doi.org/10.1111/head.14136.

Luedtke, K., May, A., 2017. Stratifying migraine patients based on dynamic pain provocation over the upper cervical spine. J. Headache Pain 18 (1), 97. https://doi.org/10.1186/s10194-017-0808-0.

Luomajoki, H.A., et al., 2018. Effectiveness of movement control exercise on patients with non-specific low back pain and movement control impairment: a systematic review and meta-analysis. Musculoskelet. Sci. Pract. 36, 1–11. https://doi.org/10.1016/j.msksp.2018.03.008.

Mingels, S., et al., 2016. Comparative analysis of head-tilt and forward head position during laptop use between females with postural induced headache and healthy controls. J. Bodyw. Mov. Ther. 20 (3), 533–541. https://doi.org/10.1016/j.jbmt.2015.11.015.

Mottram, S., et al., 2020. Assessment of movement coordination strategies to inform health of movement and guide retraining interventions. Musculoskelet. Sci. Pract. 45, 102100 https://doi.org/10.1016/j.msksp.2019.102100.

Nagaya, T., et al., 2018. Long working hours directly and indirectly (via short sleep duration) induce headache even in healthy white-collar men: cross-sectional and 1-year follow-up analyses. Int. Arch. Occup. Environ. Health 91 (1), 67–75. https://doi.org/10.1007/s00420-017-1255-6.

Patroncini, M., et al., 2014. Reliability of movement control tests on the cervical spine. BMC Muscoskel. Disord. 15 (1), 402. https://doi.org/10.1186/1471-2474-15-402.

Rendas-Baum, R., et al., 2014. Validation of the headache impact test (HIT-6) in patients with chronic migraine. Health Qual. Life Outcome 12, 117. https://doi.org/10.1186/s12955-014-0117-0.

Rodrigues, A., et al., 2021. Craniocervical flexion test in patients with migraine: discriminative validity and accuracy. Int. J. Clin. Pract. 75 (7), e14248 https://doi.org/10.1111/ijcp.14248.

Rota, E., et al., 2016. Efficacy of a workplace relaxation exercise program on muscle tenderness in a working community with headache and neck pain: a longitudinal, controlled study. Eur. J. Phys. Rehabil. Med. 52 (4), 457–465. https://pubmed.ncbi.nlm.nih.gov/26745361/.

Sahrmann, S., 2010. Movement System Impairment Syndromes of the Extremities, Cervical and Thoracic Spines. Elsevier.

Segarra, V., et al., 2015. Inter-and intra-tester reliability of a battery of cervical movement control dysfunction tests. Man. Ther. 20 (4), 570–579. https://doi.org/10.1016/j.math.2015.01.007.

Sim, J., et al., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys. Ther. 85 (3), 257–268. https://doi.org/10.1093/ptj/85.3.257.

Sleijser-Koehorst, M.L.S., et al., 2021. Diagnostic accuracy of patient interview items and clinical tests for cervical radiculopathy. Physiotherapy 111, 74–82. https://doi.org/10.1016/j.physio.2020.07.007.

Stovner, L.J., et al., 2018. Global, regional, and national burden of migraine and tension-type headache, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 17 (11), 954–976. https://doi.org/10.1016/S1474-4422(18)30322-3.

Streiner, D.L., Norman, G.R., 2008. Health Measurement Scales. Oxford University Press, Oxford, UK.

Swanenburg, J., et al., 2014. Validity and reliability of a German version of the neck disability index (NDI-G). Man. Ther. 19 (1), 52–58. https://doi.org/10.1016/j.math.2013.07.004.

Szikszay, T.M., et al., 2019. Which examination tests detect differences in cervical musculoskeletal impairments in people with migraine? A systematic review and meta-analysis. Phys. Ther. 99 (5), 549–569. https://doi.org/10.1093/ptj/pzz007.

Tegern, M., et al., 2018. Inter-rater and test-retest reliability of movement control tests for the neck, shoulder, thoracic, lumbar, and hip regions in military personnel. PLoS One 13 (9), e0204552. https://doi.org/10.1371/journal.pone.0204552.

Uthaikhup, S., et al., 2017. Effectiveness of physiotherapy for seniors with recurrent headaches associated with neck pain and dysfunction: a randomized controlled trial. Spine J. 17 (1), 46–55. https://doi.org/10.1016/j.spinee.2016.08.008.

Vos, T., et al., 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396 (10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9.

Wallwork, S.B., et al., 2020. Implicit motor imagery performance is impaired in people with chronic, but not acute, neck pain. PeerJ 8, e8553. https://doi.org/10.7717/peerj.8553.