Inter- and Intra-Rater-Reliability of a Clinical Framework for spine-related neck-arm pain

Abstract

Objective

A mechanism-based clinical framework for spine-related pain differentiates between (i) somatic referred pain, ii) heightened nerve mechanosensitivity, iii) radicular pain, iv) radiculopathy and mixed-pain. This study aimed to determine the reliability of the proposed framework.

Method

Fifty-one people with unilateral spine-related neck-arm pain were assessed and categorized by examiner 1. The classifications were compared to those made by two other examiners, based on written documentation of examiner 1. Cohens kappa was calculated between examiner-pairs; Fleiss Kappa among all examiners to assess both agreement in classifying subgroups and entire framework.

Result

Inter-rater-reliability showed moderate to almost perfect reliability (somatic: no variation, mechanosensitivity: 0.96 (95% CI 0.87-1.0) to 1.0 (95 % CI: 1.0-1.0), radicular pain: 0.46 (95% CI: 0.19 - 0.69) to 0.62 (95% CI: 0.42-0.81), radiculopathy: 0.65 (95% CI: 0.43-0.84) to 0.80 (95% CI: 0.63-0.96) mixed-pain: 0.54 (95% CI: 0.21-0.81) to 0.75 (95% CI: 0.48-0.94). There was almost perfect to moderate reliability among all examiners (somatic: no variation, mechanosensitivity: 0.97 (95% CI: 0.82–1.0), radicular pain: 0.56 (95 % CI: 0.40–0.71), radiculopathy: 0.74 (95% CI: 0.58-0.90), mixed-pain: 0.63 (95% CI: 0.47-0.79), entire framework: (0.64 (95 % CI: 0.57–0.71)).

Intra-rater-reliability showed substantial to almost perfect reliability (somatic: no variation, mechanosensitivity: 0.96 (95 % CI: 0.87–1.0), radicular pain: 0.76 (95 % CI: 0.57–0.92), radiculopathy: 0.84 (95% CI: 0.67-0.96), mixed-pain: 0.83 (95% CI: 0.60-1.0), entire framework: (0.80 (95 % CI: 0.61-0.92).

Conclusion

Substantial reliability in classifying people with spine-related neck-arm pain supports the reliability of the entire classification.

Keywords: Classification, spine-related neck-arm pain, reliability, pain type

Manuscript word count: 3483

Introduction

Neck-arm pain is heterogeneous in its presentation of clinical signs and symptoms due to various underlying pain types and mechanisms[1-3]. Presentations of mixed pain are sometimes difficult to disentangle based on localization and pain quality[4]. Classifying people into relevant clinical groups can be helpful to guide tailored treatment[5, 6]. Existing classifications for neck and neck-arm pain differ in their subgrouping criteria[7-24], and include stage of disorder (acute, sub-acute and chronic)[13-15], specific and nonspecific neck pain[12, 13], localization, episodes, pain severity and impairment[16] as well as different treatment- or mechanism-based models[17-21]. Some of these classifications describe subgroups suggesting nerve involvement, such as “radiating pain”[9, 17, 18, 22], “radicular arm pain”[23], “neurological symptoms”[24], “specific neck pain”[12, 13, 19] ,“neuropathic”[10] or different degrees of neuropathic pain[7, 25].

In recent years, mechanism-based subgrouping approaches have gained increasing interest in a range of conditions[5, 26, 27]. In the heterogeneous presentation of neck-arm pain, people may show dominantly nociceptive pain caused by activation of nociceptors in target tissue (e.g., muscles, joints, ligaments, fascia and tendons) also referred to as somatic referred pain[28, 29]. Activation of nociceptors in connective tissues surrounding nerves may cause clinical signs of heightened nerve mechanosensitivity which is considered to reflect a specific type of somatic referred pain[28, 30, 31]. Other people present with dominantly neuropathic pain, defined as pain as a direct consequence of a lesion or disease affecting the somatosensory system[29, 32, 33], as seen in radiculopathy with radicular pain[34, 35]. Radiculopathy is defined as a sensory and/or motor deficit caused by a conduction block of a spinal nerve or its nerve root and presents with clinical signs of loss of function[28, 34]. Radicular pain is evoked by ectopic discharges emanating from a dorsal root or its ganglion[28] and can present with lancinating, shock-like or electric pain[36, 37]. Whereas these presentations can occur in isolation, mixed pain presentations,

with nociceptive and/or neuropathic components are common. For instance, radicular pain can occur with radiculopathy[38] and/or with heightened nerve mechanosensitivity[2, 37, 39].

There is a lack of classifications for spine-related neck-arm pain, which distinguish between radicular pain and radiculopathy as different entities, taking into account the underlying pain mechanisms and the clinical presentation. XXX and XXX proposed a mechanism-based clinical framework for spine-related leg pain[39], which can be applied to spine-related neck-arm pain[40]. The clinical framework distinguishes between subgroups with spine-related limb pain without neurological deficits (somatic referred pain, heightened neural mechanosensitivity, radicular pain) and with neurological deficits (radiculopathy) and includes the option of mixed pain presentations[39, 40]. The reliability of this classification has not yet been investigated in people with neck-arm pain. The purpose of this study was to determine the inter- and intra-rater reliability of the proposed clinical framework[39] in people with neck-arm pain.

Methods:

Study design

This reliability study was conducted from July 2020 until November 2021 and was part of an ongoing cross-sectional study investigating somatosensory and psychosocial characteristics based on the chosen clinical framework[41]. The study protocol was approved by the Ethics Committee of the University of Applied Sciences XXX (HSOS/2019/2/2). The study protocol adheres to the ethical guidelines of the Declaration of Helsinki[42] and was funded by the University of Applied Sciences XXX and Physio Deutschland. The funders played no role in the design, conduct, or reporting of this study.

Study sample

The sample size was calculated with R Software irr package[43] and based on an expected significant and at least moderate agreement (Kappa > 0.4) between each pair of examiners and for each clinical subgroup with a significance level of 0.05 and a power of 90%. A minimum sample size of 49 people was determined. Fifty-one people were recruited from physiotherapy and medical clinics in XXX (Germany) and the surrounding area. A screening interview via telephone was conducted by examiner 1 (CK) with each potential person to verify the inclusion and exclusion criteria, using a standardized questionnaire. The inclusion criteria were unilateral spine-related neck-arm pain in people aged 18–75 years[41]. The neck-arm pain should be > 2/10 on a numeric rating scale and be provoked by movements or static positions of the cervical spine. This study was a part of a cross-sectional study investigating somatosensory profiles of people with neck-arm pain. All conditions which may confound the function of the somatosensory nervous system were excluded like: previous spinal surgery, current or previous systemic medical conditions (e.g. rheumatoid arthritis, diabetes, thyroid disease) central nervous system disorder, complex regional pain syndrome, peripheral

vascular disease, blood clotting disorder, pregnancy, psychiatric disease. As well as the presence of musculoskeletal shoulder, elbow or hand disorders in the last three months[41].

Clinical examination and classification

Procedure

The study was conducted with three physiotherapists. Two examiners (examiner 1 and 3) had a Master's degree in Orthopedic Manual Physiotherapy (OMPT) with 20 and 10 years work experience in musculoskeletal therapy, respectively. One examiner (examiner 2) was in his final year of the Master's degree (OMPT) with seven years work experience in musculoskeletal therapy.

Examiner 1 performed a comprehensive clinical examination and completed a written standardized clinical examination form (Supplement S1) containing all information required to classify people[41] (Figure 1, Table 1). This included a body chart, pain descriptors, pain behavior and symptom history. In addition, comorbidities, medications, special questions (red flags, including weight loss, spinal cord signs, vertebral artery), sleep behavior (numeric rating scale, 0 = good sleep; 10 = poor sleep) and MRI, CT or X-ray results, if available, were documented, the imaging tests were not included in the clinical decision procedure. Sex, age, pain severity on the numerical rating scale (during the last 4 weeks on average), pain duration and neck disability index were collected. All tests used in the physical examination to subgroup people have shown at least moderate reliability[44-50]. The physical examination included active movements using a Cervical Range-of-Motion device[44-46, 51]. Passive movements of the cervical spine (passive physiological and accessory intervertebral movements[49]) and the Spurling test[34, 44]. The bedside neurological examination included myotomal strength testing[47, 52], reflex testing (absent/ reduced/ normal/ increased) and sensory testing of light touch (cotton wool)[47] and pin prick sensation (neurotip) in dermatomes[53].

Neural mechanosensitivity was assessed using upper limb neurodynamic tests. ULNT 1 was performed first, followed by ULNT 2a, 2b and 3. If one test was confirmed as positive, no further ULNT tests were undertaken. Tests were considered positive when they at least partially reproduced the person's pain and the pain was altered with structural differentiation[48]. After the clinical examination, examiner 1 categorized people into the subgroups (Fig 1, Table 1).

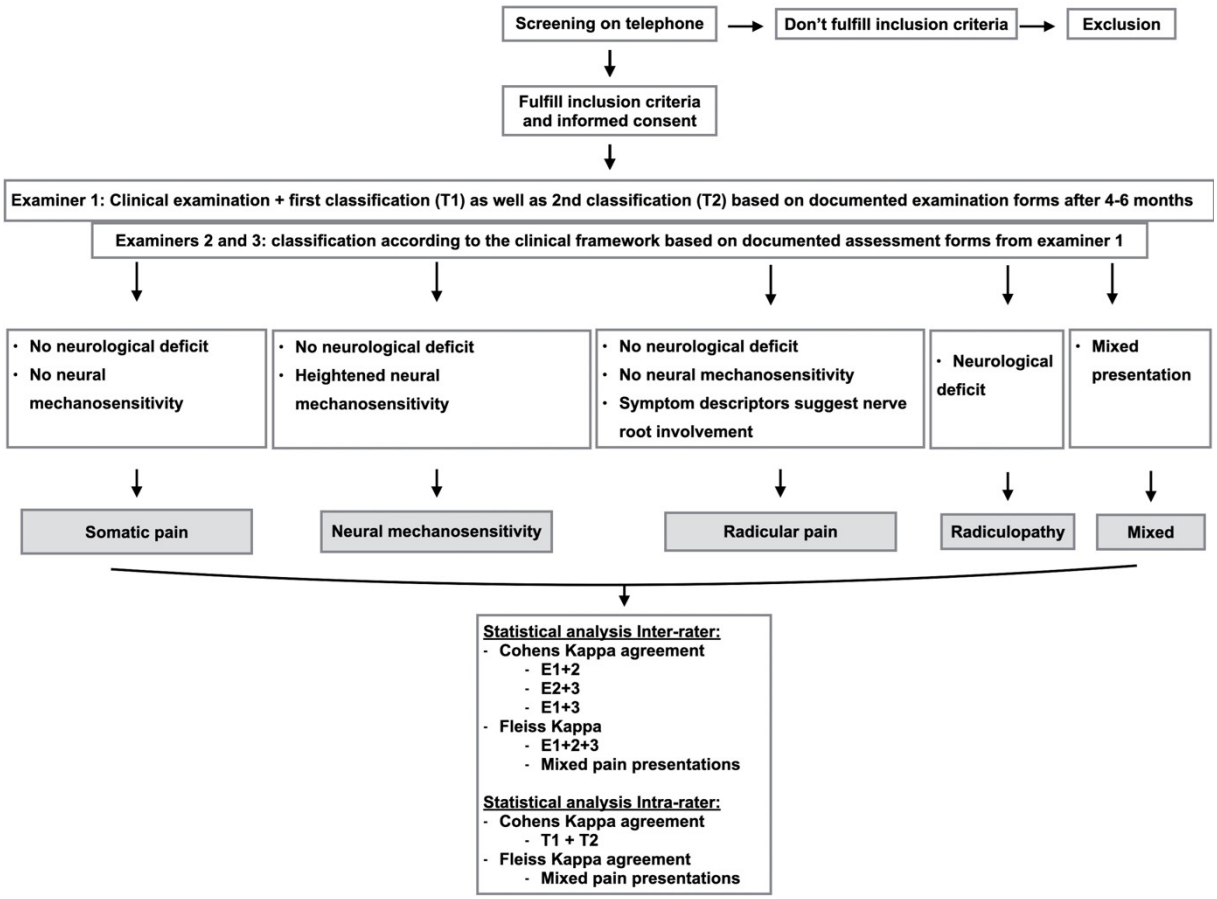

Figure 1. Flowchart study procedure. E1: examiner 1; E2: examiner 2; E3: examiner 3; T1: examiner 1, first classification; T2: examiner 1, second classification.

Table 1. Description of classification subgroups and their clinical test criteria

Subgroup and Description	Tests	Test Outcome
Somatic referred pain:	Active cervical movement	Reproduce at least

Neurological integrity and tests for neural mechanosensitivity are normal.		partially person's pain
	Bedside neurological examination (Soft touch, Pinprick, Myotomal strength, Reflex testing)	Negative
	ULNT 1, 2a, 2b, 3	Negative
	Passive cervical movement (PPIVMs and PAIVMs)	May reproduce at least partially person's pain
Neural mechanosensitivity: Neurological integrity tests are normal. Neurodynamic tests show heightened neural mechanosensitivity	Active cervical movement	Reproduce at least partially person's pain
	Bedside neurological examination (Soft touch, Pinprick, Myotomal strength, Reflex testing)	Negative
	ULNT 1, 2a, 2b, 3	Have to at least partially reproduce person's symptoms and symptoms have to alter with structural differentiation maneuvers
	Passive cervical movement (PPIVMs and PAIVMs)	May reproduce at least partially person's pain
Radicular pain:	Active cervical movement	Reproduce at least partially person's pain

The description of pain strongly suggests the involvement of a nerve root (e.g., burning, pins and needles, shooting pain, electric shock, cold pain etc.). Pain may be reported in areas reminiscent of but not necessarily identical to dermatomes. The neurological integrity tests show no loss of function and neurodynamic tests are normal. Sensory tests may demonstrate hyperalgesia/allodynia. This subgroup is based on integration of information rather than a single test or assessment cut-off.	Bedside neurological examination (Soft touch, Pinprick, Myotomal strength, Reflex testing)	Negative concerning loss of function. Sensory tests may demonstrate hyperalgesia/allodynia (gain of function).
	ULNT 1, 2a, 2b, 3	Negative
	Passive cervical movement (PPIVMs and PAIVMs)	May reproduce at least partially person's pain
Radiculopathy: Myotomal or dermatomal neuroanatomical plausible neurological deficit. Since radiculopathy is not defined by pain, it has to be grouped with one of the other clinical subgroups [39]. This subgroup is based on integration of information rather than the mere presence of loss of function. For instance, a sensory loss over the cervical spine without any other neurological signs indicative of nerve root involvement is unlikely to represent a	Active cervical movement	Not applicable
	Bedside neurological examination (Soft touch, Pinprick, Myotomal strength, Reflex testing)	Myotomal or dermatomal neuroanatomical plausible neurological deficit (loss of function)
	ULNT 1, 2a, 2b, 3	Not applicable
	Passive cervical movement (PPIVMs and PAIVMs)	Not applicable

true radiculopathy.		
Mixed pain: There is a mixed pain presentation of the different pain types with or without neurological deficits and with or without heightened neural mechanosensitivity. In this subgroup, we also recorded the specific mix of the pain presentation (e.g., radicular pain with radiculopathy) and the dominant subgroup	Active cervical movement	Reproduce at least partially person's pain
	Bedside neurological examination (Soft touch, Pinprick, Myotomal strength, Reflex testing)	Myotomal or dermatomal neuroanatomical plausible neurological deficit may occur (loss of function) Sensory tests may demonstrate hyperalgesia/allodynia (gain of function).
	ULNT 1, 2a, 2b, 3	May reproduce person's pain and pain have to alter with structural differentiation maneuvers.
	Passive cervical movement (PPIVMs and PAIVMs)	May reproduce at least partially person's pain

ULNT: Upper limb Neurodynamic Test, PPIVMs: Passive Physiological Intervertebral Movements, PAIVMs: Passive Accessory intervertebral Movements

The allocation to the groups was dichotomous (present/not present). The mixed pain presentation was assigned if people were allocated to at least two and maximum of all four

subgroups. Neck pain and limb pain were both rated separately but combined in analyses as the overall neck-arm pain presentation. For example, a person may present with somatic neck pain and with radicular arm pain.

Examiners 2 and 3 were trained by examiner 1 in three online sessions. The first online session was used to familiarize examiners with the clinical framework. In the remaining sessions, ten clinical examination forms, which were not included in the final analysis were independently rated by all examiners and then discussed.

Examiners 2 and 3 received a copy of all clinical examination notes. They were blinded towards the classifications by examiner 1 and classified people independently according to the clinical framework, based on the written clinical examination forms and their own clinical reasoning skills. We deliberately chose this approach to assess reliability of the clinical reasoning process rather than reliability associated with specific assessments, which has been reported to be moderate to substantial elsewhere[44-50].

For intra-rater-reliability, examiner 1 subgrouped all people immediately after the clinical examination (timepoint 1 = T1). To minimize a potential memory effect[54], examiner 1 re-classified all people again after 4 - 6 months (timepoint 2 = T2) based on the clinical examination notes. For the re-classification, examiner 1 was blinded to all information which may have led to the recognition of the individual person and their classification.

Analysis:

The statistical analysis was performed by examiner 1. All data were pseudonymized. Statistical analyses were carried out with SPSS Version 27. Mean and standard deviations of people's demographic and clinical data were calculated.

Cohen's kappa and absolute agreement (percentage agreement without chance adjustment) were calculated to assess the reliability between two examiners (examiner 1 and 2, examiner 1 and 3, examiner 2 and 3) and Fleiss kappa to assess the reliability between all examiners in classifying people into subgroups (table 2).

As a mixed neck-arm pain presentation can include two or up to four groups per person, it was important to determine if the examiners agreed in the specific presentations within the mixed pain subgroup, which is referred to as the reliability of the clinical framework. To determine the reliability for the clinical framework among examiners**,** Fleiss kappa was used (table 3).

To evaluate the reliability of the dominant subgroup of each person Cohens kappa (between examiner), absolute agreement and Fleiss kappa (between all examiners) were calculated (Supplement S2).

For intra-rater-reliability Cohen's kappa and absolute agreement were calculated for the agreement classifying into each subgroup. Cohens kappa was determined for the mixed pain presentations. A kappa value of > 0.00 shows poor reliability. Kappa values between 0 to 0.20 indicate slight reliability, values from 0.21 to 0.40 are considered fair, kappa between 0.41 to 0.60 indicate moderate reliability, values from 0.61 to 0.80 indicate substantial, and values from 0.81 to 1 indicate almost perfect reliability[55]. Kappa values which are at least moderate were considered acceptable in this study[54].

Results:

Fifty-one people with neck-arm pain participated in this study (15 males; 36 females). The mean age was 42,9 years (SD: 11,6). The mean pain intensity on the numerical rating scale (during the last 4 weeks on average) was 4,8 (SD: 1,9) and the mean pain duration 196 weeks (SD: 266,8). The neck disability index showed a mean of 13,5 (SD: 6,5) out of 50 possible points, indicating mild disability[56]. MRIs were available for 27 people. In eight participants, the MRI showed no abnormalities; 19 had signs of nerve root compression at C5-7, caused by a disc protrusion, herniated disc or foraminal stenosis.

Figure 2 shows the frequencies of each clinical group classified by every examiner. All three examiners determined that all participants had somatic pain (exclusively or as part of mixed pain) in their neck-arm pain presentation. Because of the mixed pain presentation, multiple counts of subgroups were possible.

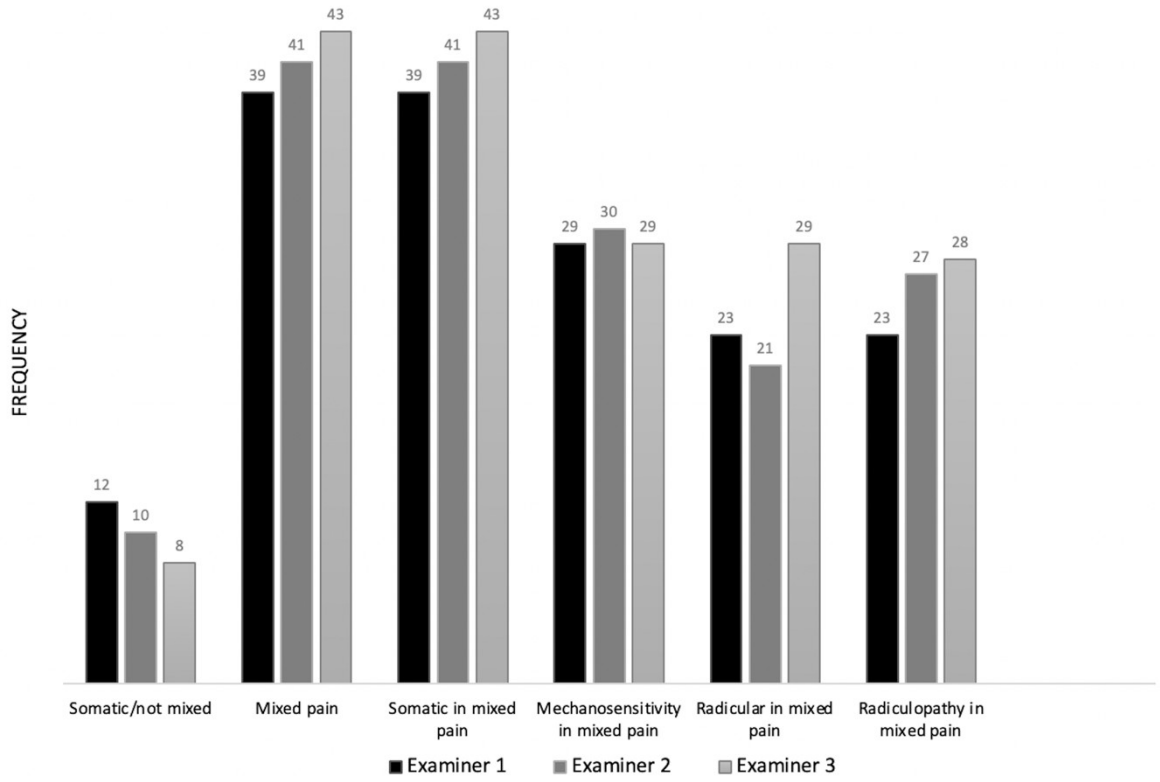

Figure 2. Frequencies (amount) of the classified clinical subgroups by examiner. The mixed pain presentation allowed multiple mentions.

Inter-rater-reliability

Agreement between two examiners

The Cohens kappa and absolute agreement for each subgroup are shown in Table 2 (Supplement S3: cross tables). Kappa calculation for the group somatic pain was not possible due to a lack of variance in response options[57]. All examiners answered exclusively that somatic pain was *present.* The reliability between examiner pairs for all subgroups reached acceptable kappa values (at least moderate to almost perfect). The Cohens kappa for the clinical framework showed a substantial reliability between examiners 1 and 2 (K: 0.69; 95 % CI: 0.54–0.83) as well as 2 and 3 (K: 0.65; 95 % CI: 0.50–0.78). A moderate reliability for the clinical framework was detected between examiners 1 and 3 (K: 0.58; 95 % CI: 0.44–0.73).

Agreement across all examiners

The Fleiss kappa for each subgroup is demonstrated in Table 2. Kappa values indicated moderate to almost perfect reliability; the lowest reliability (moderate) was documented for the radicular pain classification. Fleiss Kappa for the clinical framework across all examiners showed a substantial reliability (K: 0.64; 95 % CI: 0.57-0.71, Table 2). Table 3 presents the seven specific presentations within the mixed pain subgroup and the Fleiss Kappa values. The reliability for the dominant subgroup showed moderate reliability between all examiners and examiner pairs (Supplement S2).

Table 2. Pairwise Cohens Kappa and 95 % confidence intervals (CI) and percentage agreement for each clinical subgroup are documented for each examiner pair. Fleiss Kappa and 95 % CI are reported for every clinical subgroup across all three examiners.

Examiner	Cohens Kappa	95 % CI	% Agreement	Fleiss Kappa	95% CI
Somatic pain					
E 1 – E 2	*	*	100		
E 1 – E 3	*	*	100	*	*
E 2 – E 3	*	*	100		
Heightened neural mechanosensitivity					
E 1 – E 2	0.96	0.87 - 1.00	98		
E 1 – E 3	1.00	1.00 - 1.00	100	0.97	0.82 - 1.00
E 2 – E 3	0.96	0.86 - 1.00	98		
Radicular pain					
E 1 – E 2	0.60	0.37 - 0.80	80		
E 1 – E 3	0.46	0.19 - 0.69	73	0.56	0.40 - 0.71
E 2 – E 3	0.62	0.42 - 0.81	80		
Radiculopathy					
E 1 – E 2	0.77	0.59 - 0.92	88		
E 1 – E 3	0.65	0.43 - 0.84	82	0.74	0.58 - 0.90
E 2 – E 3	0.80	0.63 - 0.96	90		
Mixed pain					
E 1 – E 2	0.54	0.21 - 0.81	84		
E 1 – E 3	0.75	0.48 - 0.94	92	0.63	0.47 - 0.79
E 2 – E 3	0.60	0.25 - 0.85	88		
Clinical Framework					
E 1 – E 2	0.69	0.54 – 0.83	73		
E 1 – E 3	0.58	0.44 – 0.73	65	0.64	0.57 – 0.71

E 2 – E 3	0.65	0.50 – 0.78	71

* Cohens Kappa was not calculable due to lack of variance in responses between examiners.

Table 3. Number of people classified by each examiner (E1-3) into the respective subgroups and the associated Fleiss Kappa with confidence intervals (95% CI).

Somatic	**Mechano-sensitivity**	**Radicular pain**	**Radiculo-pathy**	**E1**	**E2**	**E3**	**Fleiss Kappa (CI)**
x				13	10	8	0.80 (0.64 – 0.96)
x	x			12	11	7	0.71 (0.55 – 0.87)
x	x	x		1	1	6	0.21 (0.05 – 0.37)
x	x	x	x	13	12	10	0.70 (0.55 – 0.86)
x		x		2	2	2	0.48 (0.32 – 0.64)
x			x	0	3	1	0.23 (0.07 – 0.39)
x		x	x	7	6	11	0.61 (0.45 – 0.76)
x	x		x	3	6	6	0.56 (0.40 – 0.72)
Total				51	51	51	0.64 (0.57 – 0.71)

E1: Examiner 1; E2: Examiner 2; E3: Examiner 3.

Intra-rater-agreement:

The Cohens Kappa for each single clinical subgroup is demonstrated in Table 4 (Supplement S3: cross tables). Cohens Kappa showed almost perfect reliability for all subgroups between the timepoints T**1** und T**2**. Cohens Kappa for the clinical framework showed a substantial reliability between timepoints (K: 0.80; 95% CI: 0.61-0.92). Table 5 presents the specific presentations within the mixed pain subgroup and the Cohens Kappa values.

Table 4. Cohens Kappa with confidence interval (95% CI) and percentage agreement between the classification at timepoint T1 and at T2.

Timepoint	Cohens Kappa	95% CI	% Agreement
Somatic pain			
T1 –T2	*	*	100
Heightened neural mechanosensitivity			
T1 –T2	0.96	0.87 - 1.00	98
Radicular pain			
T1 –T2	0.84	0.65 - 0.96	92
Radiculopathy			
T1 –T2	0.84	0.67 - 0.96	92
Mixed pain			
T1 –T2	0.83	0.60 - 1.00	94
Clinical Framework			
T1 -T2	0.80	0.61 - 0.92	84

* Cohens Kappa was not calculable due to lack of variance in responses between T1 and T2.

Table 5. Number of people classified by examiner 1 into the respective subgroups and the associated Cohens Kappa with confidence intervals (95% CI) at timepoints T1 and T2.

Somatic	**Mechano-sensitivity**	**Radicular pain**	**Radiculo-pathy**	**T0**	**T1**	**Cohens Kappa (95% CI)**
x				13	11	0.89 (0.62 - 1.00)
x	x			12	12	0.78 (0.51 - 1.00)
x	x	x		1	3	0,48 (0.21 - 0.75)
x	x	x	x	13	13	0.90 (0.62 - 1.00)
x		x		2	2	0.48 (0.21 - 0.75)
x			x	0	1	-0.10 (-0.28 - 0.27)
x		x	x	7	7	0.83 (0.56 - 1.00)
x	x		x	3	2	0.79 (0.52 - 1.00)
Total				51	51	0.80 (0.61 - 0.92)

Discussion

This is the first study investigating the reliability of the clinical framework from XXX and XXX[39] in people with spine-related neck-arm pain. There was a substantial reliability for the entire clinical framework and a high absolute agreement with moderate to almost perfect kappa values between examiners in classifying people into single subgroups. These results support the reliability of this classification system.

The high intra-rater (84-100%) as well as inter-rater (65-100%) absolute agreement in subgrouping attest to the reliability of this classification system. In particular, the subgrouping of somatic pain showed 100% absolute agreement and heightened neural mechanosensitivity showed almost perfect reliability among examiners. In comparison, Tampin et al reported moderate reliability between two examiners in classifying people with non-specific neck-arm pain associated with heightened nerve mechanosensitivity[58]. This difference may be explained by methodological differences. In the study by Tampin et al[58], both examiners performed a clinical assessment, whereas here the second ratings were based on written documentation. The comparison of inter-rater-reliability based on written documentation used here, reflects agreement in clinical reasoning and cognitive ability in subgrouping individuals based on prespecified criteria, while removing the potential bias associated with a physical examination test. Reliability thus refers to the classification process and not to the entire test procedure.

The inter-rater-reliability among all examiners for the classification of radicular pain was moderate, and for radiculopathy substantial, with more obvious discrepancies between examiner 1 and examiner 3. Some differences were seen in the interpretation of positive signs, when symptoms like paresthesia (tingling), a common descriptor of radicular pain[28, 59], were only reproduced with palpation of a muscle[60] and not with any other physical assessment suggestive of a nerve root or spinal nerve involvement[61]. In these cases, examiner 1 did not

interpret the reproduced paresthesia as a sign of radicular pain, but examiner 3 did. In some instances, the interpretation of a coherent pattern of loss of function as a sign of radiculopathy was challenging. For example, some people had only subtle neurological deficits compared to the asymptomatic arm, or sensory deficits were detected in a proximal extradermatomal area (e.g., paravertebral Th 3-6), but no distal dermatomal deficits indicative of a radiculopathy were present[37]. In the latter case, examiner 1 did not interpret the loss of function as a sign of radiculopathy, because loss of sensory function in extradermatomal areas has been reported in people with nociceptive non-specific neck–arm pain without radiculopathy[4, 62, 63]. However, examiner 3 interpreted this as a sign of radiculopathy.

This current study clearly reveals the mix of pain presentations (75-84%) in people with spine-related neck-arm pain. This heterogeneity in people with neck-arm pain has also been shown by others[58, 62, 63]. However, because the radiculopathy group is defined by neurologic deficits rather than pain, it can only occur in combination with another subgroup; this represents an increased likelihood of a mixed presentation in this study.

The identification of the exact mixed pain presentation is of high clinical relevance as management may vary between presentations[64]. For example, Schaefer et al showed that the proportion of people with low-back related leg pain responding favorably to neural mobilization techniques was higher in people with heightened nerve mechanosensitivity compared to people with somatic pain or lumbar radiculopathy[6].

The advantage of our classification system is that it allows to capture mixed pain presentations compared to other classification systems that use hierarchical decision-making[7, 58, 65]. The latter approach treats pain presentations as being mutually exclusive, which does not reflect the real world of clinical presentations of people with spine-related neck-arm pain[1, 4, 62]. The strength of our study is that our examiners were able to identify individual specific

presentations within the mixed pain subgroup as well as showing substantial reliability in establishing the exact mix of pain presentations. The reliability of the specific presentations within the mixed pain subgroup varied from poor to substantial. (Table 3 and 5). Of note, when pain presentations were combined with radiculopathy or radicular pain, they showed only fair to moderate reliability. It seems that particularly the classification of radicular pain was more challenging than others, as also seen earlier in the inter-examiner comparison (Table 2). This may be explained by the absence of agreed diagnostic criteria and resulting vaguer criteria for radicular pain which largely rely on the integration of patterns of subjective symptoms rather than objective signs as seen in radiculopathy (loss of function) or heightened neural mechanosensitivity (positive neurodynamic tests). As a matter of fact, it was left to the clinicians' clinical reasoning skills how to interpret these symptom patterns. These observations suggest that there is a need for further specification of clinical criteria and assessment findings indicative of the presence of radicular pain and radiculopathy.

One mixed pain presentation in the intra-rater-agreement showed a negative kappa value (somatic pain with radiculopathy). This is due to the fact that this pain presentation was not classified at the first measurement and only one time at the second measurement. Since this is only one pain presentation out of 51, statistical evaluation and interpretation is limited.

Schaefer et al[65, 66] developed a classification system for low back related leg pain, which has also been applied in people with non-specific arm pain[7]. The system defines four subgroups (central sensitization, denervation, peripheral nerve sensitization, musculoskeletal) which, to some extent, are comparable to our subgroups. Their denervation group reflects radiculopathy, peripheral nerve sensitization is equivalent to heightened nerve mechanosensitivity and the musculoskeletal group to somatic pain. Reliability studies on the Schaefer classification revealed a substantial inter-rater-reliability between two examiners[7, 65].

One study compared the agreement based on clinical examination (Kappa 0.72; CI: 0.57-0.86)[65], the other study based on written documentation (kappa: 0.78; no CI calculated)[7]. The results similarly showed substantial reliability with only slightly higher kappa values compared to ours. This may be explained by methodological differences. A hierarchical classification is likely easier to apply than identifying mixed pain presentations. Moloney et al classified 40 people with non-specific neck-arm pain with the Schaefer classification and the Smart classification[7], which allows classifying into a mixed pain group, (composed of the following subgroups: nociceptive pain, neuropathic pain and central sensitization). The reliability was moderate. They concluded that the mixed pain presentation makes the clinical decision making more difficult but reflects the real-life clinical situation[7].

This study investigated both intra-and inter-rater-reliability in our study as both aspects are important. While sufficient inter-rater-reliability is important to guarantee that different clinicians agree in clinical decision making, sufficient intra-rater-reliability is a basis and crucial to correctly monitor a person's clinical course over time. A clinician must be confident that a clinical presentation which changes over the course of treatment is an actual change and not the result of misclassification due to low intra-rater-reliability.

Study limitations

The main limitation of the study pertains to the fact that examiner 2 and 3 did not clinically assess the people. This was a pragmatic and conscious decision, as this study was mainly interested in the reliability of clinical reasoning and the process of classification rather than clinical tests. Indeed, reliability of tests used to subgroup patients such as upper limb neurodynamic tests, bedside neurological examination or assessing active and passive movements has already been reported to be moderate to substantial[44-48, 50]. This suggests that the performance of assessments is unlikely a major confounder of reliability of these

classification[7, 65]. The method of using written documentation to establish inter-rater-reliability has been well documented[7, 58, 67]. One advantage of this methodical approach is that examiners are given the same information for their clinical judgement whereby clinical examination may pose the risk to obtain variable test results due to repeated measurements or a change in the health condition[54]. Also, repeated assessment would have added a considerable burden to people and might have exacerbated the people's condition thus challenging interpretation. Further, it should be noted that the data collection fell entirely within the Covid-19 pandemic, during which it was essential to avoid unnecessary human contact. It remains to be shown whether agreement remains at high levels if different therapists perform separate examinations.

The examiners in this study had undergone postgraduate training and 7-20 years working experience. It is unclear if less experienced therapists would achieve similar outcomes. In addition, the documentation of the imaging tests may have interfered the examiner's decision, although it is not a part of the clinical tests necessary for the clinical framework.

The sample size was calculated for the agreement between two examiners in classifying people into the five different subgroups, but not for the classification of specific pain presentations within the mixed pain subgroup. The analyses for the latter were underpowered, hence results have to be interpreted with caution.

Conclusion

There was a substantial inter-rater and intra-rater-reliability for the entire clinical framework and a high absolute agreement between examiners in classifying people into the single subgroups. Clearer definitions for the clinical presentation of radicular pain would be helpful to improve classification of this group. The clinical framework is favored over other classifications as it allows to capture mixed pain presentations, reflecting the real world of clinical presentation of people with spine-related neck-arm pain. This work together with

future studies will provide important information on the clinical usefulness of mechanism-based subgrouping to improve management outcome. Further investigation of the validity of this classification system is necessary.

References:

1. Moloney N, Hall T, Doody C. Sensory hyperalgesia is characteristic of nonspecific arm pain: a comparison with cervical radiculopathy and pain-free controls. *Clin J Pain.* Nov 2013;29(11):948-956.

2. Tampin B, Slater H, Briffa NK. Neuropathic pain components are common in patients with painful cervical radiculopathy, but not in patients with nonspecific neck-arm pain. *Clin J Pain.* Oct 2013;29(10):846-856.

3. Mitchell Tim BD, Slater Helen, O'Sullivan Peter. *Musculoskeletal Clinical Translation Framework: From Knowing to Doing*2017.

4. Tampin B, Briffa NK, Goucke R, Slater H. Identification of neuropathic pain in patients with neck/upper limb pain: application of a grading system and screening tools. *Pain.* Dec 2013;154(12):2813-2822.

5. Jull G. FD, Treleaven J., O'leary S., Lewis J.S. *Management of Neck Pain Disorders: a Research-Informed Appoach.* Vol 1: Elsevier; 2019.

6. Schafer A, Hall T, Muller G, Briffa K. Outcomes differ between subgroups of patients with low back and leg pain following neural manual therapy: a prospective cohort study. *Eur Spine J.* Mar 2011;20(3):482-490.

7. Moloney NA, Hall TM, Leaver AM, Doody CM. The clinical utility of pain classification in non-specific arm pain. *Man Ther.* Feb 2015;20(1):157-165.

8. Rasmussen H, Kent P, Kjaer P, Kongsted A. In a secondary care setting, differences between neck pain subgroups classified using the Quebec task force classification system were typically small - a longitudinal study. *BMC Musculoskelet Disord.* Jun 16 2015;16:150.

9. Childs JD, Cleland JA, Elliott JM, et al. Neck pain: Clinical practice guidelines linked to the International Classification of Functioning, Disability, and Health from the Orthopedic Section of the American Physical Therapy Association. *J Orthop Sports Phys Ther.* Sep 2008;38(9):A1-a34.

10. Liu R, Kurihara C, Tsai HT, et al. Classification and Treatment of Chronic Neck Pain: A Longitudinal Cohort Study. *Reg Anesth Pain Med.* Jan/Feb 2017;42(1):52-61.

11. Dewitte V, Peersman W, Danneels L, Bouche K, Roets A, Cagnie B. Subjective and clinical assessment criteria suggestive for five clinical patterns discernible in nonspecific neck pain patients. A Delphi-survey of clinical experts. *Man Ther.* Dec 2016;26:87-96.

12. Huisstede BM, Miedema HS, Verhagen AP, Koes BW, Verhaar JA. Multidisciplinary consensus on the terminology and classification of complaints of the arm, neck and/or shoulder. *Occup Environ Med.* May 2007;64(5):313-319.

13. Scherer M. CJ-F. Die Körperliche Untersuchung bei Nackenschmerzen. *Zeitschrift für Allgemeinmedizin.* 2007;83(12):507 - 521.

14. Hoy DG, Protani M, De R, Buchbinder R. The epidemiology of neck pain. *Best Pract Res Clin Rheumatol.* Dec 2010;24(6):783-792.

15. Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for neck pain. *Phys Ther.* Oct 2001;81(10):1701-1717.

16. MacDermid JC, Walton DM, Bobos P, Lomotan M, Carlesso L. A Qualitative Description of Chronic Neck Pain has Implications for Outcome Assessment and Classification. *Open Orthop J.* 2016;10:746-756.

17. Fritz JM, Brennan GP. Preliminary examination of a proposed treatment-based classification system for patients receiving physical therapy interventions for neck pain. *Phys Ther.* May 2007;87(5):513-524.

18. Blanpied PR, Gross AR, Elliott JM, et al. Neck Pain: Revision 2017. *J Orthop Sports Phys Ther.* Jul 2017;47(7):A1-A83.

19. Luomajoki H. EM. Nackenschmerzen. Befund, Unterteilung und Therapie. *Physioactive.* 2017;1(2017):7-14.

20. Maissan F, Pool J, de Raaij E, Wittink H, Ostelo R. Treatment based classification systems for patients with non-specific neck pain. A systematic review. *Musculoskelet Sci Pract.* Jun 2020;47:102133.

21. Kent P, Mjosund HL, Petersen DH. Does targeting manual therapy and/or exercise improve patient outcomes in nonspecific low back pain? A systematic review. *BMC Med.* Apr 8 2010;8:22.

22. Childs JD, Fritz JM, Piva SR, Whitman JM. Proposal of a classification system for patients with neck pain. *J Orthop Sports Phys Ther.* Nov 2004;34(11):686-696; discussion 697-700.

23. Wang WT, Olson SL, Campbell AH, Hanten WP, Gleeson PB. Effectiveness of physical therapy for patients with neck pain: an individualized approach using a clinical decision-making algorithm. *Am J Phys Med Rehabil.* Mar 2003;82(3):203-218; quiz 219-221.

24. Guzman J, Hurwitz EL, Carroll LJ, et al. A new conceptual model of neck pain: linking onset, course, and care: the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976).* Feb 15 2008;33(4 Suppl):S14-23.

25. Smart KM, Blake C, Staines A, Doody C. Clinical indicators of 'nociceptive', 'peripheral neuropathic' and 'central' mechanisms of musculoskeletal pain. A Delphi survey of expert clinicians. *Man Ther.* Feb 2010;15(1):80-87.

26. Baron R, Förster M, Binder A. Subgrouping of patients with neuropathic pain according to pain-related sensory abnormalities: a first step to a stratified treatment approach. *Lancet Neurol.* Nov 2012;11(11):999-1005.

27. Vardeh D, Mannion RJ, Woolf CJ. Toward a Mechanism-Based Approach to Pain Diagnosis. *J Pain.* Sep 2016;17(9 Suppl):T50-69.

28. Bogduk N. On the definitions and physiology of back pain, referred pain, and radicular pain. *Pain.* Dec 15 2009;147(1-3):17-19.

29. Armstrong SA, Herr MJ. Physiology, Nociception. *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2022.

30. Elvey RL. Physical evaluation of the peripheral nervous system in disorders of pain and dysfunction. *J Hand Ther.* Apr-Jun 1997;10(2):122-129.

31. Sauer SK, Bove GM, Averbeck B, Reeh PW. Rat peripheral nerve components release calcitonin gene-related peptide and prostaglandin E2 in response to noxious stimuli: evidence that nervi nervorum are nociceptors. *Neuroscience.* 1999;92(1):319-325.

32. Finnerup NB, Haroutounian S, Kamerman P, et al. Neuropathic pain: an updated grading system for research and clinical practice. *Pain.* Aug 2016;157(8):1599-1606.

33. Treede RD. The International Association for the Study of Pain definition of pain: as valid in 2018 as in 1979, but in need of regularly updated footnotes. *Pain Rep.* Mar 2018;3(2):e643.

34. Lam KN, Rushton A, Thoomes E, Thoomes-de Graaf M, Heneghan NR, Falla D. Neck pain with radiculopathy: A systematic review of classification systems. *Musculoskelet Sci Pract.* Aug 2021;54:102389.

35. Baron R, Binder A, Attal N, Casale R, Dickenson AH, Treede RD. Neuropathic low back pain in clinical practice. *Eur J Pain.* Jul 2016;20(6):861-873.

36. Tampin B, Slater H, Briffa NK. Neuropathic pain components are common in patients with painful cervical radiculopathy, but not in patients with nonspecific neck-arm pain. *Clin J Pain.* 2013;20(10):846-856.

37. Tampin B, Slater H, Hall T, Lee G, Briffa NK. Quantitative sensory testing somatosensory profiles in patients with cervical radiculopathy are distinct from those in patients with nonspecific neck-arm pain. *Pain.* Dec 2012;153(12):2403-2414.

38. Tampin B, Lind C, Jacques A, Slater H. Disentangling 'sciatica' to understand and characterise somatosensory profiles and potential pain mechanisms. *Scand J Pain.* Jan 27 2022;22(1):48-58.

39. Schmidt A, Tampin B. . Spinally referred back and leg pain. Scott D. Boden (Ed): Online book of The International Society for the Study of the Lumbar Spine2018.

40. Tampin B. Nervenschmerz ≠ Nervenschmerz. *Manuelle Therapie.* 2020;Vol. 1 Februar 2020:21-27.

41. Kapitza C, Lüdtke K, Tampin B, Ballenberger N. Application and utility of a clinical framework for spinally referred neck-arm pain: A cross-sectional and longitudinal study protocol. *PLoS One.* 2020;15(12):e0244137.

42. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama.* Nov 27 2013;310(20):2191-2194.

43. Matthias Gamer JLaIFPS. Various Coefficients of Interrater Reliability and Agreement. 2019-01-26 17:07:15 UTC 2019.

44. Lemeunier N, Jeoun EB, Suri M, et al. Reliability and validity of clinical tests to assess posture, pain location, and cervical spine mobility in adults with neck pain and its associated disorders: Part 4. A systematic review from the cervical assessment and diagnosis research evaluation (CADRE) collaboration. *Musculoskelet Sci Pract.* Dec 2018;38:128-147.

45. Fletcher JP, Bandy WD. Intrarater reliability of CROM measurement of cervical spine active range of motion in persons with and without neck pain. *J Orthop Sports Phys Ther.* Oct 2008;38(10):640-645.

46. Williams MA, McCarthy CJ, Chorti A, Cooke MW, Gates S. A systematic review of reliability and validity studies of methods for measuring active and passive cervical range of motion. *J Manipulative Physiol Ther.* Feb 2010;33(2):138-155.

47. Schmid AB, Brunner F, Luomajoki H, et al. Reliability of clinical tests to evaluate nerve function and mechanosensitivity of the upper limb peripheral nervous system. *BMC Musculoskelet Disord.* Jan 21 2009;10:11.

48. Nee RJ, Jull GA, Vicenzino B, Coppieters MW. The validity of upper-limb neurodynamic tests for detecting peripheral neuropathic pain. *J Orthop Sports Phys Ther.* May 2012;42(5):413-424.

49. Hall T, Briffa K, Hopper D, Robinson K. Reliability of manual examination and frequency of symptomatic cervical motion segment dysfunction in cervicogenic headache. *Man Ther.* Dec 2010;15(6):542-546.

50. Lemeunier N, da Silva-Oolup S, Chow N, et al. Reliability and validity of clinical tests to assess the anatomical integrity of the cervical spine in adults with neck pain and its associated disorders: Part 1-A systematic review from the Cervical Assessment and Diagnosis Research Evaluation (CADRE) Collaboration. *Eur Spine J.* Sep 2017;26(9):2225-2241.

51. Thoomes-de Graaf M, Thoomes E, Fernández-de-Las-Peñas C, Plaza-Manzano G, Cleland JA. Normative values of cervical range of motion for both children and adults: A systematic review. *Musculoskelet Sci Pract.* Oct 2020;49:102182.

52. Kendall FP ME, Provance PG. Muscles Testing and Function *4th edition. Baltimore: Williams and Wilkins.* 1993.

53. Ridehalgh C, Sandy-Hindmarch OP, Schmid AB. Validity of Clinical Small-Fiber Sensory Testing to Detect Small-Nerve Fiber Degeneration. *J Orthop Sports Phys Ther.* Oct 2018;48(10):767-774.

54. Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. *Qual Life Res.* Aug 2014;23(6):1713-1720.

55. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* Mar 1977;33(1):159-174.

56. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther.* Sep 1991;14(7):409-415.

57. Stucky BD, Pereira CCA. Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol: Measurement in medicine: a practical guide. *Quality of Life Research.* 2012;21(2):371-373.

58. Tampin B, Briffa NK, Hall T, Lee G, Slater H. Inter-therapist agreement in classifying patients with cervical radiculopathy and patients with non-specific neck-arm pain. *Man Ther.* Oct 2012;17(5):445-450.

59. Smith SM, McMullen CW, Herring SA. Differential Diagnosis for the Painful Tingling Arm. *Curr Sports Med Rep.* Sep 1 2021;20(9):462-469.

60. Kwon J, Kim HS, Chang WH, Park C, Lee SC. Characteristics of Myofascial Pain Syndrome of the Infraspinatus Muscle. *Ann Rehabil Med.* Aug 2017;41(4):573-581.

61. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A, Allison S. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine (Phila Pa 1976).* Jan 1 2003;28(1):52-62.

62. Moloney N, Hall T, Doody C. Divergent sensory phenotypes in nonspecific arm pain: comparisons with cervical radiculopathy. *Arch Phys Med Rehabil.* Feb 2015;96(2):269-275.

63. Ottiger-Boettger K, Ballenberger N, Landmann G, Stockinger L, Tampin B, Schmid A. Somatosensory profiles in patients with non-specific neck-arm pain with and without positive neurodynamic tests. *Musculoskelet Sci Pract.* Sep 17 2020;50:102261.

64. Freynhagen R, Parada HA, Calderon-Ospina CA, et al. Current understanding of the mixed pain concept: a brief narrative review. *Curr Med Res Opin.* Jun 2019;35(6):1011-1018.

65. Schäfer A, Hall TM, Lüdtke K, Mallwitz J, Briffa NK. Interrater reliability of a new classification system for patients with neural low back-related leg pain. *J Man Manip Ther.* 2009;17(2):109-117.

66. Schafer A, Hall T, Briffa K. Classification of low back-related leg pain--a proposed patho-mechanism-based approach. *Man Ther.* Apr 2009;14(2):222-230.

67. Tampin B, Broe RE, Seow LL, et al. Field testing of the revised neuropathic pain grading system in a cohort of patients with neck and upper limb pain. *Scand J Pain.* Jul 26 2019;19(3):523-532.