

Clinical pain research

Brigitte Tampin*, Rachel Elizabeth Broe^a, Lee Lee Seow^a, Shushana Gijohn George^a, Jijie Tan^a, Rajiv Menon, Angela Jacques and Helen Slater

Field testing of the revised neuropathic pain grading system in a cohort of patients with neck and upper limb pain

<https://doi.org/10.1515/sjpain-2018-0348>

Received November 30, 2018; revised February 15, 2019; accepted February 18, 2019; previously published online March 22, 2019

Abstract

Background and aims: In 2008, the International Association for the Study of Pain Special Interest Group on Neuropathic Pain (NeuPSIG) proposed a clinical grading system to help identify patients with neuropathic pain (NeP). We previously applied this classification system, along with two NeP screening tools, the painDETECT (PD-Q) and Leeds Assessment of Neuropathic Symptoms and Signs pain scale (LANSS), to identify NeP in patients with neck/upper limb pain. Both screening tools failed to identify a large proportion of patients with clinically classified NeP, however a limitation of our study was the use of a single clinician performing the NeP classification. In 2016, the NeuPSIG grading system was updated with the

aim of improving its clinical utility. We were interested in field testing of the revised grading system, in particular in the application of the grading system and the agreement of interpretation of clinical findings. The primary aim of the current study was to explore the application of the NeuPSIG revised grading system based on patient records and to establish the inter-rater agreement of detecting NeP. A secondary aim was to investigate the level of agreement in detecting NeP between the revised NeuPSIG grading system and the LANSS and PD-Q.

Methods: In this retrospective study, two expert clinicians (Specialist Pain Medicine Physician and Advanced Scope Physiotherapist) independently reviewed 152 patient case notes and classified them according to the revised grading system. The consensus of the expert clinicians' clinical classification was used as "gold standard" to determine the diagnostic accuracy of the two NeP screening tools.

Results: The two clinicians agreed in classifying 117 out of 152 patients (ICC 0.794, 95% CI 0.716–850; κ 0.62, 95% CI 0.50–0.73), yielding a 77% agreement. Compared to the clinicians' consensus, both LANSS and PD-Q demonstrated limited diagnostic accuracy in detecting NeP (LANSS sensitivity 24%, specificity 97%; PD-Q sensitivity 53%, specificity 67%).

Conclusions: The application of the revised NeP grading system was feasible in our retrospective analysis of patients with neck/upper limb pain. High inter-rater percentage agreement was demonstrated. The hierarchical order of classification may lead to false negative classification. We propose that in the absence of sensory changes or diagnostic tests in patients with neck/upper limb pain, classification of NeP may be further improved using a cluster of clinical findings that confirm a relevant nerve lesion/disease, such as reflex and motor changes. The diagnostic accuracy of LANSS and PD-Q in identifying NeP in patients with neck/upper limb pain remains limited. Clinical judgment remains crucial to diagnosing NeP in the clinical practice.

^aRachel Elizabeth Broe, Lee Lee Seow, Shushana Gijohn George and Jijie Tan: These authors contributed equally to this work.

*Corresponding author: Brigitte Tampin, Department of Physiotherapy, Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands, WA, Australia; Department of Neurosurgery, Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands, WA, Australia; School of Physiotherapy and Exercise Science, Curtin University, Perth, WA, Australia; and Faculty of Business Management and Social Sciences, Hochschule Osnabrück, University of Applied Sciences, Osnabrück, Germany,

Phone: +61 8 6457 7964, Fax: + 61 8 6457 3481,

E-mail: Brigitte.Tampin@health.wa.gov.au

Rachel Elizabeth Broe, Lee Lee Seow, Shushana Gijohn George, Jijie Tan and Helen Slater: School of Physiotherapy and Exercise Science, Curtin University, Perth, WA, Australia

Rajiv Menon: Department of Pain Management, Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands, WA, Australia

Angela Jacques: School of Physiotherapy and Exercise Science, Curtin University, Perth, WA, Australia; and Department of Research, Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands, WA, Australia

Implications: Our observations suggest that in view of the heterogeneity in patients with neck/upper limb pain, a considerable amount of expertise is required to interpret the revised grading system. While the application was feasible in our clinical setting, it is unclear if this will be feasible to apply in primary health care settings where early recognition and timely intervention is often most needed. The use of LANSS and PD-Q in the identification of NeP in patients with neck/upper limb pain remains questionable.

Keywords: neuropathic pain; validation; grading system; screening questionnaires.

1 Introduction

In order to improve health outcomes, people with neuropathic pain (NeP) should be identified early and receive timely evidence-based intervention and targeted treatments [1, 2]. Patients with NeP, defined as “pain caused by a lesion or disease of the somatosensory nervous system” [3], exhibit higher pain intensity and more prolonged suffering than those with nociceptive pain [4]. They require longer duration of treatment [4], and have a lower quality of life. Not surprisingly, failure to implement appropriate treatments for NeP is associated with poorer health outcomes, and imposes increased health care costs and strain on limited health resources [4]. Leading an important reform initiative in 2008, Treede and colleagues proposed a clinical grading system to assist in identifying patients

with NeP [5]. This system classified NeP into four levels of certainty – no, possible, probable and definite NeP, based on the four criteria as shown in Fig. 1 [5].

This 2008 grading system has been widely applied in different patient cohorts [7–10]. A recent review has highlighted various difficulties with the application of this system in real world clinical settings [6]. These include an uncertainty about determining a clear causal relationship between pain and a neural lesion and also determining the location and underlying pathology of a neural lesion [6, 8]. In response to such identified issues and to better align with clinical practice, a refinement in the order of the diagnostic criteria and improved guidance about application of the grading system was proposed by Finnerup and colleagues (Fig. 1) [6].

We have previously applied the original classification system, along with two NeP screening tools, the pain-DETECT (PD-Q) and Leeds Assessment of Neuropathic Symptoms and Signs pain scale (LANSS), to identify NeP in patients with neck/upper limb pain [9]. We found that both screening tools failed to identify a large proportion of patients with clinically classified definite NeP [9]. A limitation of our previous study was also the use of a single clinician performing the classification, thus increasing the risk of diagnostic bias.

Field testing of the NeuPSIG revised grading system may involve clinical patient assessment by two independent clinicians to explore inter-rater agreement, however this adds considerable responder burden for patients and clinicians. In our current study we were interested

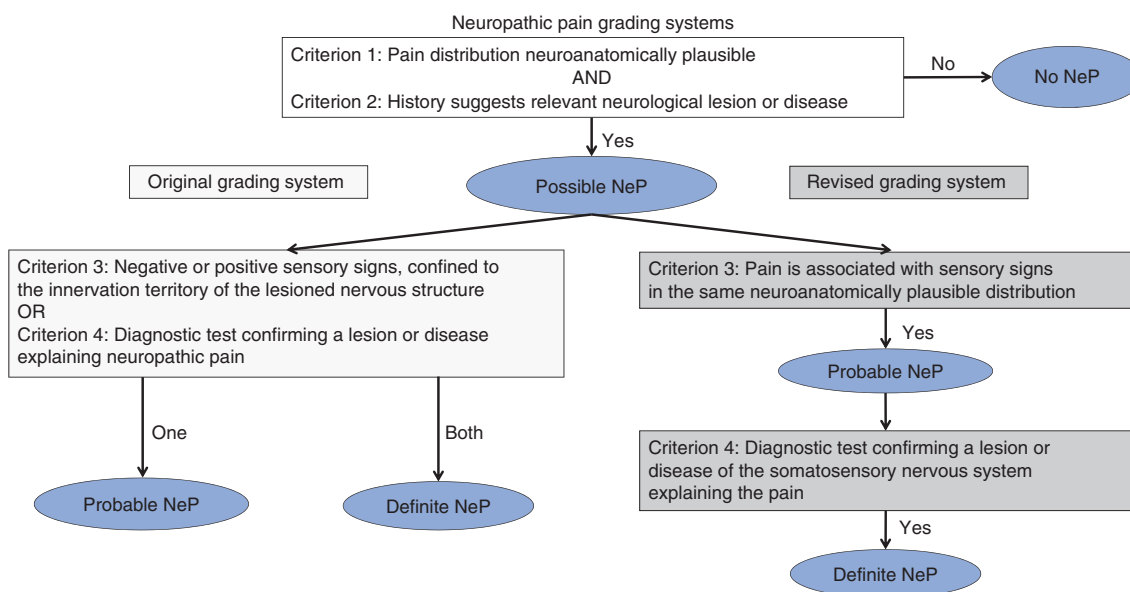


Fig. 1: Flowchart showing the classification of neuropathic pain (NeP) according to the original grading system [5] (left; light grey) and according to the revised grading system [6] (right; darker grey).

in the second step, the interpretation of clinical findings and application of the grading system based on medical records as this may be a useful method to enhance clinicians' diagnostic skills in the assessment of NeP, as well as a clinical audit. Numerous studies used this methodology to determine the level of certainty of NeP based on the original grading system [6]. Clinical record audits are valid tools to assess and ensure quality and safety in health services. The primary aim of our study was to explore the application of the NeuPSIG revised grading system based on patient records and to establish the inter-rater agreement of detecting NeP. A secondary aim was to investigate the level of agreement in detecting NeP between the revised NeuPSIG grading system and the LANSS and PD-Q.

2 Methods

This was a retrospective study on 152 patients with neck and upper limb pain with suspected nerve lesion attending a neurosurgery outpatient clinic in a large tertiary hospital [9]. The current study was registered with the Quality Improvement Unit of Sir Charles Gairdner Hospital (registration number 14430) and endorsed by the Hospital's Human Research Ethics Committee on 18 April 2017 and Curtin University Human Research Ethics Committee (HREC 2017-0505). The study protocol adhered to the Declaration of Helsinki.

2.1 Classification protocol

A detailed description of the patient cohort, previous clinical assessments and methodology is described in Tampin et al. [9]. Patients in the previous study had been recruited consecutively. The clinical assessment included the patient's history, pain drawings, documentation of pain descriptors and pain behaviors, musculoskeletal assessments and neurological examination. Sensory testing comprised the assessment of light touch and pinprick sensation in both upper limbs for determination of dermatomal sensory deficits and in both lower limbs, if spinal cord compromise was suspected, as well as in the maximal pain area for the assessment of NeP. Sensory testing findings were compared with the contralateral corresponding control site or, in case of bilateral pain, in proximal or distal pain-free sites [11]. Assessment of thermal sensitivity was not performed, consistent with previously documented methodologies [12–15].

For the purpose of this study, the patients' clinical notes were de-identified and electronically scanned by four postgraduate research students (JT, RB, LS, SG). Two highly experienced clinicians working in the field of NeP (RM, BT) independently reviewed the patient notes and imaging reports, a Pain Medicine Consultant (RM with 2 years of Consultant Practice) and an Advanced Scope Physiotherapist (BT; 9 years of Advanced Scope Practice and 26 years of musculoskeletal postgraduate experience) who had performed the initial patient assessment 8 years ago [9]. The latter had no recollection of the previous patient classification, however as an additional step to reduce bias, patient records were de-identified.

Prior to the independent clinical audit and reclassification, these two clinicians developed a recording sheet for the application of the revised NeP grading system [6] (Supplementary Table 1). Several randomly chosen imaging reports were reviewed in order to establish a consensus on which information would constitute a confirmatory test (Supplementary Table 2). To avoid false positives, diagnostic images were interpreted conservatively. Consistent with the previous study, only radiologists' reports stating a compromise of the nerve root at the clinically-relevant level were deemed as a confirmatory test [9]. A pilot audit of 10 case notes was conducted prior to the full audit to allow consensus building between reviewers regarding the application of these criteria. A list of abbreviations used in the patient notes was compiled in order to allow clarity of interpretation of patient assessments (Supplementary Table 3). The 10 case notes and imaging reports in our pilot study were drawn from the patient sample used to establish consensus between clinicians. This method complied with recommended research strategies and consensus setting for retrospective audits [16].

Each clinician then independently reviewed the de-identified patient clinical notes, completed the recording sheets and classified the patients according to the revised NeP grading system. Audits were completed in batches of approximately 20–30 files at a time to ensure consistency. The first 20 completed cases were subsequently also re-evaluated by each reviewer at the end of the audit to ensure consistency. The clinicians were blinded from PD-Q and LANSS scores while classifying patients. Each clinician's recording sheets were collected in a sealed envelope to maintain blinding prior to data analysis. The experts' clinical diagnosis was then entered into an excel spreadsheet by the research students. Data entry was crosschecked among these students to ensure accuracy. This spreadsheet was merged with previously captured data on the two screening tools, PD-Q and LANSS questionnaires.

2.2 Clinical screening tools

2.2.1 LANSS

The LANSS discriminates between patients with or without pain of predominantly neuropathic origin and is applied in an interview format [12]. The LANSS contains five sensory descriptor items and two clinical examination items. LANSS was developed in 60 patients with distinct clinical diagnostic categories of NeP and non-NeP, and demonstrated a sensitivity of 83% and specificity of 87%, and was further validated in 40 patients (sensitivity 85%, specificity of 80%) [12]. A score of <12 indicates NeP being unlikely to contribute to a patient's symptoms and a score ≥ 12 as being highly likely neuropathic in origin [12].

2.2.2 PainDETECT

The PD-Q is a self-administered tool consisting of seven weighted sensory descriptors, plus one item relating to spatial pain characteristics and one relating to temporal characteristics [4]. The questionnaire classifies patients into three groups: a NeP component is unlikely, results are ambiguous, or a NeP component is likely. The PD-Q original validation study undertaken in 392 patients with clinically diagnosed pain of predominantly either nociceptive or neuropathic origin demonstrated a sensitivity of 85% and specificity of 80% [4]. A PD-Q score ≤ 12 signifies that a NeP component is unlikely and a score ≥ 19 indicates a likely NeP component. The PD-Q was designed to identify NeP components specifically in low back pain patients with and without referred leg pain. The questionnaire had not been validated for neck pain patients with or without referred arm pain prior to our previous study [9].

In our previous study [9], the participants completed the PD-Q before they attended the clinical examination and the clinician was blinded to these results. The LANSS was administered in an interview format at the end of the clinical examination as it required clinical intervention in the form of bedside testing for allodynia and pinprick threshold recording.

2.3 Statistical analysis

The Statistical Package for Social Sciences (SPSS version 24.0, IBM, Armonk, NY, USA) was used to analyse the data. The demographic and clinical pain characteristics

of patients were compared between four pain groups; no NeP, possible NeP, probable NeP and definite NeP, using Kruskal-Wallis and a one-way analysis of variance. Pairwise comparison of the following was performed:

- Inter-rater agreement in classifying NeP according to the revised NeP grading system.
- Clinical classification based on clinicians' consensus and LANSS
- Clinical classification based on clinicians' consensus and PD-Q

The percentage of agreement, Cohen's κ coefficient, the intraclass-correlation coefficient (ICC) and 95% CIs were calculated to determine the degree of concordance between the two clinicians in classifying the four categories of NeP (no, possible, probable, definite). The κ /ICC strength of agreement is classified as: >0 = less than chance agreement, 0.01 – 0.20 = poor agreement, between 0.21 and 0.40 = fair agreement, between 0.41 and 0.60 = moderate agreement, 0.61 – 0.80 = substantial agreement and 0.81 – 1.00 = almost perfect agreement [17].

A secondary analysis was performed to investigate the agreement in neuropathic pain classifications between the original and revised grading systems. Cohen's κ with 95% CIs and the percentage of agreement were calculated.

To align with the dichotomous scale used in LANSS, dichotomous variables were applied for PD-Q (<19 = unlikely NeP and ≥ 19 = likely NeP) and the clinical classification (no/possible = unlikely NeP; and probable/definite = NeP). The percentage of agreement, Cohen's κ coefficient and 95% CI were calculated for the agreement between clinical classification and screening tools in classifying unlikely NeP and NeP.

The sensitivity and specificity of LANSS and PD-Q were calculated using the consensus of the expert clinicians' clinical classification based on the revised grading system applied as the "gold standard". Predictive values, odds and likelihood ratios were also calculated. The receiver operating characteristics (ROC) and the area under the curve (AUC) with their 95% confidence was computed for each screening questionnaire. A calculated probability of $p < 0.05$ was considered significant for all analyses.

3 Results

The study was conducted between April 2017 and October 2017. One hundred fifty-two patients' clinical notes were reviewed. Details of the patient characteristics of the total cohort are available in Tampin et al. [9].

3.1 Patient characteristics

Out of the 117 patients meeting clinical consensus, none were classified as having no NeP (Tables 1 and 2). Thirty-nine patients (33.3%) were classified as having possible NeP, a large proportion of patients were classified as having probable NeP ($n=66$, 56.4%), while 12 patients (10.3%) were classified as having definite NeP (Tables 1 and 2). Table 2 illustrates patient demographics.

3.2 Inter-rater agreement on neuropathic pain classification

The two examiners agreed in classifying 117 out of 152 patients (ICC 0.794, 95% CI 0.716–850; κ 0.62, 95% CI 0.50–0.73) (Table 1), yielding a 77% overall agreement. The number of patients in each grouping based on the clinical classifications by the two independent clinicians is presented in Table 1. Out of the 117 patients, diagnostic testing results (such as imaging reports) were

not available in 13 (11%) patients. Six of these patients were classified as having possible NeP, five patients as having probable NeP. The remaining two patients were classified by one clinician as having possible, and by the other clinician as having probable NeP, and *vice versa*. No patient with only localized neck pain, and an absence of sensory findings, was classified as having probable NeP.

In 35 cases (23.0%), no consensus was reached between clinicians ($\chi^2=30.63$, $p<0.001$) (Table 3). Diagnostic testing results were not available for two patients in the non-consensus group. There was a disparity in NeP classification group between both clinicians in 19 cases (16.2%) (Table 3). Out of the 19 cases, eight patients were classified under the “NeP” group by clinician 1 (RM), while these patients were classified under the “unlikely NeP” group by clinician 2 (BT). The other 11 cases were classified by clinician 1 (RM) under the “unlikely NeP” group and by clinician 2 (BT) under the “NeP” group.

Table 1: Inter-rater clinical consensus for each neuropathic pain classification group ($n=152$).

	Clinician 1 (RM)				Total
	No NeP	Possible NeP	Probable NeP	Definite NeP	
Clinician 2 (BT)					
No NeP	0	3	1	1	5
Possible NeP	4	39	6	0	49
Probable NeP	1	9	66	3	79
Definite NeP	0	1	6	12	19
Total	5	52	79	16	152

NeP=neuropathic pain.

Table 3: No clinical consensus between clinicians ($n=35$).

Clinician 2	Clinician 1				Total
	Unlikely NeP		NeP		
	No NeP	Possible NeP	Probable NeP	Definite NeP	
Unlikely NeP					
No NeP	0	3	1	1	5
Possible NeP	4	0	6	0	10
NeP					
Probable NeP	1	9	0	3	13
Definite NeP	0	1	6	0	7
Total	5	13	13	4	35

NeP=neuropathic pain.

Table 2: Patient characteristics ($n=117$) based on clinicians' consensus.

	Total patients	Possible NeP	Probable NeP	Definite NeP
n	117	39 (26%)	66 (43%)	12 (8%)
Age (years) ^a	50.6±11.6	49.9±11.8	51.3±11.2	49.3±13.1
Gender (male/female)	59/58	21/18	30/36	8/4
Symptoms duration (months) ^b	12.0 (6.0–24.0)	12.0 (6.0–27.0)	12.0 (6.8–24.0)	8.5 (2.5–17.0)
Pain now (NRS 0–10) ^a	4.8±2.2	4.7±2.5	4.9±2.1	4.5±2.1
Maximal pain intensity during last 4 weeks ^a	7.7±2.1	7.4±2.2	7.9±2.0	7.0±2.6
Average pain intensity during last 4 weeks ^a	6.0±2.0	6.5±2.1	6.0±1.9	4.9±2.1
painDETECT score	17.2±6.6	16.5±6.0	17.7±7.1	16.8±6.3
LANSS score	8.0 (3.5–11.0)	5.0 (1.0–8.0)	10.0 (6.0–12.0)	8.5 (5.8–10.0)

NeP=neuropathic pain; NRS=numeric rating scale; LANSS=Leeds Assessment of Neuropathic Symptoms and Signs pain scale. ^aMean±SD.

^bMedian (interquartile range).

Most of the non-consensus cases related to the classification between possible and probable NeP. In nine cases, clinician 1 (RM) classified the patients as having possible NeP, while clinician 2 (BT) classified them as having probable NeP. Six cases were classified by the clinician 2 (BT) as having possible NeP and were categorized as having probable NeP by clinician 1 (RM).

3.2.1 Agreement on neuropathic pain classifications between the original and revised grading systems

The original and revised clinical classification systems differed in agreement across the four levels of NeP classification ($\chi^2=56.54$, $p<0.001$). Of the 117 patients, 64 cases (54.7%) were re-classified into the same NeP groups (no NeP=0; possible NeP=18; probable NeP=36; definite NeP=10) (Supplementary Table 4), yielding a fair agreement of 54.7% between both grading systems in detecting NeP (κ 0.31, 95% CI 0.18–0.45). Compared to the original grading system, the application of the revised grading system resulted in the identification of a greater number of patients in the possible and probable NeP groups, and less numbers in the definite NeP group (Supplementary Table 4).

3.3 Agreement in neuropathic pain classification between the clinicians' consensus and questionnaires where patients were classified as having NeP or no NeP

3.3.1 LANSS and the clinicians' consensus

NeP classification differed between the LANSS and the clinicians' consensus ($\chi^2=8.72$, $p=0.002$). LANSS identified 20 patients (17.1%) with NeP [mean 14.0 (range 12.0–17.0)] and 97 cases (82.9%) with unlikely NeP [mean 7.0 (range 0.0–11.0)] (Table 4). There was poor agreement between LANSS and the clinicians' consensus in 57 out of the 117 cases (unlikely NeP, $n=38$; NeP, $n=19$; κ 0.159, CI 0.082–0.260), yielding 49% of agreement with 24.4% sensitivity and 97.4% specificity (Table 5). The predictive values, odds and likelihood ratios are listed in Table 5. Of the remaining 60 cases, 59 (98.3%) were classified as having unlikely NeP according to LANSS while clinicians' consensus classified them as having NeP. The AUC curve for LANSS was 0.76 (95% CI 0.68–0.85; $p<0.001$)

Table 4: Neuropathic pain (NeP) classification based on clinicians' consensus, LANSS and painDETECT.

	Clinicians' consensus		Total
	Unlikely NeP ^a	Likely NeP	
LANSS ^b			
Unlikely NeP	38	59	97
Likely NeP	1	19	20
Total	39	78	117
painDETECT ^c			
Unlikely NeP	26	37	63
NeP	13	41	54
Total	39	78	117

^aNo and possible NeP grouped as Unlikely NeP. ^b49% agreement between clinical classification and LANSS. ^c57% agreement between clinical classification and painDETECT.

Table 5: Accuracy of screening questionnaires in identifying patients with neuropathic pain.

	% Sensitivity	% Specificity	PPV	NPV	LR+	LR–	DOR
LANSS	24.4	97.4	0.95	0.39	9.38	0.78	12.0
PD-Q	52.6	66.7	0.76	0.41	1.58	0.71	2.2

PPV = positive predictive value; NPV = negative predictive value; LR = likelihood ratio; DOR = diagnostic odds ratio; LANSS = Leeds Assessment of Neuropathic Symptoms and Signs pain scale; PD-Q = painDETECT questionnaire.

and the appropriate cutoff score for our population was 7.5 (Fig. 2).

3.3.2 PD-Q and the clinicians' consensus

NeP classification differed between the PD-Q and the clinicians' consensus ($\chi^2=3.87$, $p=0.038$). PD-Q identified 54 patients (46.2%) with NeP [mean 23.0, (range 19.6–23.4)] and 63 cases (53.8%) with unlikely NeP [mean 12.2, (range 8.1–16.3)] (Table 4). Poor agreement was demonstrated between PD-Q and the clinicians' consensus in 67 out of the 117 cases (unlikely NeP group, $n=26$; NeP group, $n=41$; κ 0.17, 95% CI: –0.00–0.33). Using the clinicians' consensus as the “gold standard”, PD-Q yielded a 57% agreement with a sensitivity of 52.6% and specificity of 66.7% (Table 5). Of the remaining 50 cases, 37 (74.0%) were classified as having unlikely NeP according to PD-Q compared to the clinicians' consensus as having NeP. The AUC curve for PD-Q was 0.56 (95% CI 0.45–0.66; $p=0.330$) with an appropriate cut-off score for our population of 17.5 (Fig. 2).

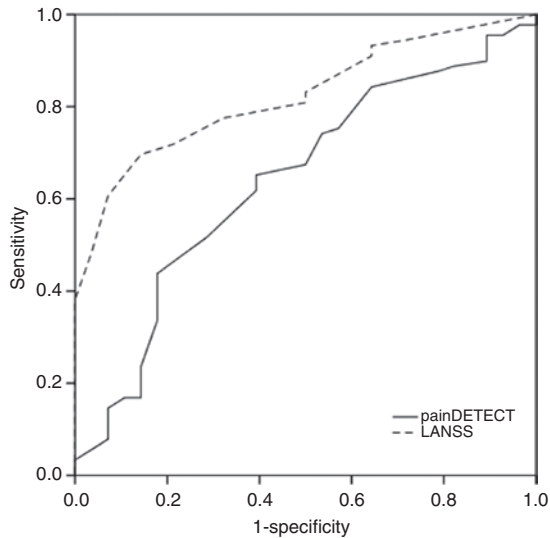


Fig. 2: Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the Leeds Assessment of Neuropathic Symptoms and Signs pain scale (LANSS) and painDETECT questionnaire based on the revised grading system.

4 Discussion

4.1 Main findings

To our knowledge, this is the first study to field test the clinical application of the revised NeP grading system in patients with neck and upper limb pain. A high inter-rater percentage agreement was demonstrated in classifying NeP. LANSS and PD-Q failed to detect a large proportion of patients with NeP.

The high inter-rater percentage agreement (77%) demonstrated in NeP classification indicates that the revised NeP grading system can be used reliably between skilled clinicians. The inter-rater agreement justified the use of the clinician's consensus as the "gold standard" for a comparison with NeP screening tools and therefore enhanced the validity of the clinical NeP classification in this study compared to our previous study. While the percentage agreement between examiners was high, the κ value of 0.62 was just above the cut-off for substantial strength agreement, and the lower range of the confidence interval was within the moderate agreement range [17]. The κ coefficient can be a conservative estimate of agreement [18, 19] as this takes into account agreement occurring by chance. Given the application of a robust diagnostic work-up here, it is unlikely these classifications were based on guesswork. The ICC of 0.794 (95% CI 0.716–850) indicated substantial

agreement between examiners and was close to almost perfect agreement.

Consensus was not reached between clinicians in 35 cases. In 16 of these cases (46%), patients were still classified by both clinicians into the same group: either "unlikely NeP" (no/possible) or the "NeP" group (probable/definite NeP) and therefore, would arguably have obtained the same clinical management [6]. Therefore, the disagreement was not necessarily clinically critical, as treatment recommendations would have been similar. For the remaining 19 cases (54%), discrepancies reflected different classifications into the "unlikely NeP" group or the "NeP" group. This is clinically important, as treatment approaches would differ according to these NeP classifications [6].

Reasons for disparity in classification were mostly due to differences between clinicians in the interpretation of sensory changes and imaging reports. Some patients had no evidence of sensory changes in their main pain area, however imaging reports were interpreted by one clinician as a confirmatory test and yet not by the other. The revised grading system proposes a hierarchical order of classification, hence in the absence of sensory changes the classification of probable NeP cannot be achieved as one cannot progress to the next step as this criterion has not been satisfied [6]. However, the authors of the revised grading system do comment in the legend of their Fig. 2, that in some cases sensory changes may be difficult to demonstrate, although the nature of the lesion is confirmed. For these cases, the classification "probable" would still apply if a diagnostic test confirmed a lesion [6]. While our clinicians did take these aspects into account, they still differed in their interpretation of some imaging reports. We propose that classification may be further improved where other clinical findings such as reflex and motor changes at relevant nerve root levels may be considered as a cluster of clinical findings that confirm a relevant lesion/disease.

Some imaging reports stated the presence of severe canal or foraminal stenosis, but did not comment on the presence or absence of neural compromise. In these cases, if the clinical presentation and clinical examination findings were consistent with a neural lesion, the imaging report could be interpreted as a confirmatory diagnostic test. In six cases, no consensus was met due to such ambiguity in interpreting patient clinical notes. These issues reflect the real world, as primary care clinicians rely on radiologist's reports and a large proportion of imaging findings may be false positives [20] and possibly not directly relevant to treatment decisions.

Our observations suggest that in view of the heterogeneity in patients with neck and upper limb pain, a considerable amount of expertise is required to interpret the revised grading system. While the application was feasible in our clinical setting, it is unclear if this will be feasible to apply in primary health care settings where early recognition and timely intervention is often most needed.

4.2 Advantages and challenges of the revised grading system

The authors of the revised grading system have provided clearer guidance for clinicians on how to reach each level of certainty for the presence of NeP; i.e. taking into account pain descriptors for the classification of possible NeP and specifying the type of sensory change (negative sign) indicative of a nerve lesion. While the hierarchical order of classification may have clinical utility for non-specialists, this approach may lead to a false negative classification and potentially a lack of appropriate treatment. For example, a patient presenting with a C6 radicular pain distribution of symptoms and NeP pain descriptors, absent biceps reflex and C6 myotomal weakness, but without any sensory changes in the main pain area would be classified as having possible NeP. According to NeuPSIG guidelines pharmacological treatment for NeP would not be commenced [6] even though symptoms of NeP and signs of a nerve root lesion/radiculopathy are apparent. While experienced clinicians would not rely solely on such a classification system to make clinical decisions, this is not the case for primary care settings where decision support is often most needed.

4.3 Comparison of clinicians' consensus and screening questionnaires

Compared to the "gold standard", LANSS and PD-Q failed to detect a large proportion of patients with NeP, similar to results in our previous study [9]. Previously, LANSS demonstrated a sensitivity of 22% and specificity of 88% while PD-Q demonstrated a sensitivity of 64% and specificity of 62% [9]. The low sensitivity of LANSS (24%) could be attributed to the differing patient characteristics in our cohort compared to the cohort in the original validation study [9]. The specificity of LANSS was high, consistent with other studies [21, 22], indicating its usefulness in negating the presence of NeP in patients with neck-arm pain. The PD-Q demonstrated a sensitivity of 53% which was lower compared to our previous study and others [9, 23], and a specificity of 67% which had improved by 5% compared to previous findings [9]. The specificity may have improved

as a larger proportion of patients were clinically classified as having unlikely NeP ($n=39$) compared to our previous study ($n=28$). Nevertheless, the discriminative ability of PD-Q in identifying NeP remains questionable.

It is unclear why LANSS performed better than PD-Q with respect to specificity. A comparison of the two screening tools is complicated by the fact that they differ in their design, their pain descriptor items and the scoring of these items. It is possible that the clinical examination tests of LANSS, which corresponded with the sensory testing used by the clinician, improved specificity, but not sensitivity.

Both screening questionnaires demonstrated insufficient diagnostic accuracy, consistent with findings in other studies [7, 24, 25]. This observation is further reflected in the lowered overall agreement between the screening tools and clinical classification here compared to our previous study (LANSS 49% vs. 68%; PD-Q 57% vs. 63%). Clinicians should be cautious about relying too heavily on these tools to detect the presence of NeP. Screening tools can supplement clinical reasoning, but cannot replace it.

4.4 Strengths and limitations

The strength of our study lies in the robust methods, using two independent clinicians' rating, blinded application of the grading system and blinded data entry and analysis. A limitation is that not both clinicians examined the patients explaining some interpretation differences, although this occurred in only six cases. While a study design would be strengthened by both clinicians examining the patients, this adds considerable responder burden for patients and clinicians. We acknowledge the limitation associated with using the statistical assumption that the four NeP grades are mutually exclusive, and the distinct separation between possible and probable NeP into "unlikely NeP" and NeP. However, this is the design of the revised grading system and we applied this approach as a pragmatic decision. This highlights that in the real world, rather than the standalone application of a "grading system", clinical decision making is optimally guided by all relevant clinical findings to ensure best practice care.

4.5 Conclusion

In conclusion, the application of the revised NeP grading system was feasible in our retrospective analysis of patients with neck/upper limb pain. High inter-rater percentage agreement was demonstrated. The feasibility

and reliability of the grading system in primary health care settings needs further explorations and strengthening in decision support to ensure “right care”. The diagnostic accuracy of LANSS and PD-Q in identifying NeP in patients with neck/upper limb pain remains limited. Clinical judgment remains crucial to diagnosing NeP in the clinical practice.

Acknowledgments: This study was supported by a Grant (2017-18/033) from the Research Advisory Committee, Sir Charles Gairdner Osborne Park Health Care Group. The authors thank Dr Jenny Lalor for her statistical advice and all participants in this research.

Authors' statements

Research funding: This work was supported by a Grant (2017-18/033) from the Research Advisory Committee, Sir Charles Gairdner Osborne Park Health Care Group, Perth, Western Australia.

Conflict of interest: All authors declare no conflicts of interest.

Informed consent: Not applicable, patient records were de-identified.

Ethical approval: The study was registered with the Quality Improvement Unit of Sir Charles Gairdner Hospital (registration number 14430) and endorsed by the Hospital's Human Research Ethics Committee and Curtin University Human Research Ethics Committee (HREC 2017-0505). The study protocol adhered to the Declaration of Helsinki.

References

- [1] Finnerup NB, Attal N, Haroutounian S, McNicol E, Baron R, Dworkin RH, Gilron I, Haanpää M, Hansson P, Jensen TS, Kamerman PR, Lund K, Moore A, Raja SN, Rice ASC, Rowbotham M, Sena E, Siddall P, Smith BH, Wallace M. Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis. *Lancet Neurol* 2015;14:162–73.
- [2] Harden N, Cohen M. Unmet needs in the management of neuropathic pain. *J Pain Symptom Manage* 2003;25:S12–7.
- [3] Jensen TS, Baron R, Haanpää M, Kalso E, Loeser JD, Rice ASC, Treede R-D. A new definition of neuropathic pain. *Pain* 2011;152:2204–5.
- [4] Freynhagen R, Baron R, Gockel U, Tölle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin* 2006;22:1911–20.
- [5] Treede R-D, Jensen TS, Campbell JN, Cruccu G, Dostrovsky JO, Griffin JW, Hansson P, Hughes R, Numikko T, Serra J. Neuropathic pain. Redefinition and a grading system for clinical and research purposes. *Neurology* 2008;70:1630–5.
- [6] Finnerup NB, Haroutounian S, Kamerman P, Baron R, Bennett DLH, Bouhassira D, Cruccu G, Freeman R, Hansson P, Nurmikko T, Raja SN, Rice ASC, Serra J, Smith BH, Treede R-D, Jensen TS. Neuropathic pain: an updated grading system for research and clinical practice. *Pain* 2016;157:1599–606.
- [7] Haroutounian S, Nikolajsen L, Finnerup NB, Jensen TS. The neuropathic component in persistent postsurgical pain: a systematic literature review. *Pain* 2013;154:95–102.
- [8] Mulvey MR, Rolke R, Klepstad P, Caraceni A, Fallon M, Colvin L, Laird B, Bennett MI. Confirming neuropathic pain in cancer patients: applying the NeuPSIG grading system in clinical practice and clinical research. *Pain®* 2014;155:859–63.
- [9] 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* 2013;154:2813–22.
- [10] Themistocleous AC, Ramirez JD, Shillo PR, Lees JG, Selvarajah D, Orengo C, Tesfaye S, Rice ASC, Bennett DLH. The Pain in Neuropathy Study (PiNS): a cross-sectional observational study determining the somatosensory phenotype of painful and painless diabetic neuropathy. *Pain* 2016;157:1132–45.
- [11] Haanpää M, Attal N, Backonja M, Baron R, Bennett M, Bouhassira D, Cruccu G, Hansson P, Haythornthwaite JA, Iannetti GD, Jensen TS, Kauppila T, Nurmikko TJ, Rice ASC, Rowbotham M, Serra J, Sommer C, Smith BH, Treede R-D. NeuPSIG guidelines on neuropathic pain assessment. *Pain* 2011;152:14–27.
- [12] Bennett M. The LANSS pain scale: The Leeds Assessment of Neuropathic Symptoms and Signs. *Pain* 2001;92:147–57.
- [13] Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, Cunin G, Fermanian J, Ginies P, Grun-Overdyking A, Jafari-Schluep H, Lantéri-Minet M, Laurent B, Mick G, Serrie A, Valade D, Vicaut E. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005;114:29–36.
- [14] Haroun OMO, Hietaharju A, Bizuneh E, Tesfaye F, Brandsma JW, Haanpää M, Rice ASC, Lockwood DNJ. Investigation of neuropathic pain in treated leprosy patients in Ethiopia: a cross-sectional study. *Pain* 2012;153:1620–4.
- [15] Weingarten TN, Watson JC, Hooten WM, Wollan PC, Melton III LJ, Lockett AJ, Wong GY, Yawn BP. Validation of the S-LANSS in the community setting. *Pain* 2007;132:189–94.
- [16] Gregory KE, Radovinsky L. Research strategies that result in optimal data collection from the patient medical record. *Appl Nurs Res* 2012;25:108–16.
- [17] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [18] Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–68.
- [19] McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–82.
- [20] Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen L, Deyo R, Halabi S, Turner J, Avins A, James K. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *Am J Neuroradiol* 2015;36:811–6.
- [21] Unal-Cevik I, Sarioglu-Ay S, Evcik D. A comparison of the DN4 and LANSS questionnaires in the assessment of neuropathic pain: validity and reliability of the Turkish version of DN4. *J Pain* 2010;11:1129–35.
- [22] Yucel A, Senocak M, Kocasoy Orhan E, Cimen A, Ertas M. Results of the leeds assessment of neuropathic symptoms and signs pain scale in Turkey: a validation study. *J Pain* 2004;5:427–32.

- [23] De Andrés J, Pérez-Cajaraville J, Lopez-Alarcón MD, López-Millán JM, Margarit C, Rodrigo-Royo MD, Franco-Gay ML, Abellón D, Ruiz MA, López-Gomez V, Pérez M. Cultural adaptation and validation of the painDETECT scale into Spanish. *Clin J Pain* 2012;28:243–53.
- [24] Mulvey MR, Bennett MI, Liwowsky I, Freynhagen R. The role of screening tools in diagnosing neuropathic pain. *Pain Manag* 2014;4:233–43.
- [25] Hasvik E, Haugen AJ, Gjerstad J, Grøvlø L. Assessing neuropathic pain in patients with low back-related leg pain: comparing the painDETECT Questionnaire with the 2016 NeuPSIG grading system. *Eur J Pain* 2018;22:1160–9.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/sjpain-2018-0348>).