The reliability of clinical judgments and criteria associated with mechanisms-based classifications of pain in patients with low back pain disorders: A preliminary reliability stud...

Article *in* The Journal of manual & manipulative therapy · June 2010

DOI: 10.1179/106698110X12640740712897 · Source: PubMed

5 authors, including:

Keith M Smart
St. Vincents University Hospital
22 PUBLICATIONS **279** CITATIONS
SEE PROFILE

Catherine Blake
University College Dublin
115 PUBLICATIONS **1,079** CITATIONS
SEE PROFILE

Anthony Staines
Dublin City University
233 PUBLICATIONS **4,036** CITATIONS
SEE PROFILE

Catherine Doody
University College Dublin
48 PUBLICATIONS **639** CITATIONS
SEE PROFILE

The reliability of clinical judgments and criteria associated with mechanisms-based classifications of pain in patients with low back pain disorders: a preliminary reliability study

Keith M. Smart[1,2], Antoinette Curley[3], Catherine Blake[1], Anthony Staines[4], Catherine Doody[1]

[1]UCD School of Public Health, Physiotherapy and Population Science, University College Dublin, Ireland, [2]St Vincent's University Hospital, Ireland, [3]Physiotherapy Department, Adelaide and Meath Hospital, Ireland, [4]School of Nursing, Dublin City University, Ireland

Mechanisms-based classifications of pain have been advocated for their potential to aid understanding of clinical presentations of pain and improve clinical outcomes. However, the reliability of mechanisms-based classifications of pain and the clinical criteria upon which such classifications are based are not known. The purpose of this investigation was to assess the inter- and intra-examiner reliability of clinical judgments associated with: (i) mechanisms-based classifications of pain; and (ii) the identification and interpretation of individual symptoms and signs from a Delphi-derived expert consensus list of clinical criteria associated with mechanisms-based classifications of pain in patients with low back ($\pm$leg) pain disorders. The inter- and intra-examiner reliability of an examination protocol performed by two physiotherapists on two separate cohorts of 40 patients was assessed. Data were analysed using kappa and percentage of agreement values. Inter- and intra-examiner agreement associated with clinicians' mechanisms-based classifications of low back ($\pm$leg) pain was 'substantial' (kappa =0.77; 95% confidence interval (CI): 0.57–0.96; % agreement =87.5) and 'almost perfect' (kappa =0.96; 95% CI: 0.92–1.00; % agreement=92.5), respectively. Sixty-eight and 95% of items on the clinical criteria checklist demonstrated clinically acceptable (kappa $\geqslant$ 0.61 or % agreement $\geqslant$ 80%) inter- and intra-examiner reliability, respectively. The results of this study provide preliminary evidence supporting the reliability of clinical judgments associated with mechanisms-based classifications of pain in patients with low back ($\pm$leg) pain disorders. The reliability of mechanisms-based classifications of pain should be investigated using larger samples of patients and multiple independent examiners.

Keywords: Classification, Low back pain, Pain mechanisms, Reliability

In response to recent advances within the pain sciences, mechanisms-based classifications of pain have been advocated as a means of identifying clinically relevant sub-groups of patients who likely require differing intervention strategies in order to improve clinical outcomes.[1–3] 'Mechanisms-based classification' refers to the classification of clinical presentations of pain based on assumptions concerning the dominant underlying neurophysiological mechanisms responsible for its generation and/or maintenance.[4,5]

Correspondence to: K M Smart, St Vincent's University Hospital, Elm Park, Dublin 4, Ireland. Email: k.smart@ucd.ie

It has been argued that mechanisms-based approaches may better explain observed variations in the nature and severity of many clinical presentations of musculoskeletal pain (e.g. low back ($\pm$leg) pain, whiplash associated disorder): (i) where pain is reported in the absence of or disproportionate to any clearly identifiable pathology; (ii) where pain is reported to persist after the resolution of injury or pathology; (iii) where the severity of pain reported by patients with similar injuries and pathologies differs greatly; and paradoxically (iv) where pain does not exist despite evidence of injury or pathology.[2,6] In addition, proponents argue that mechanisms-based approaches could optimise clinical outcomes by

DOI 10.1179/106698110X12640740712897

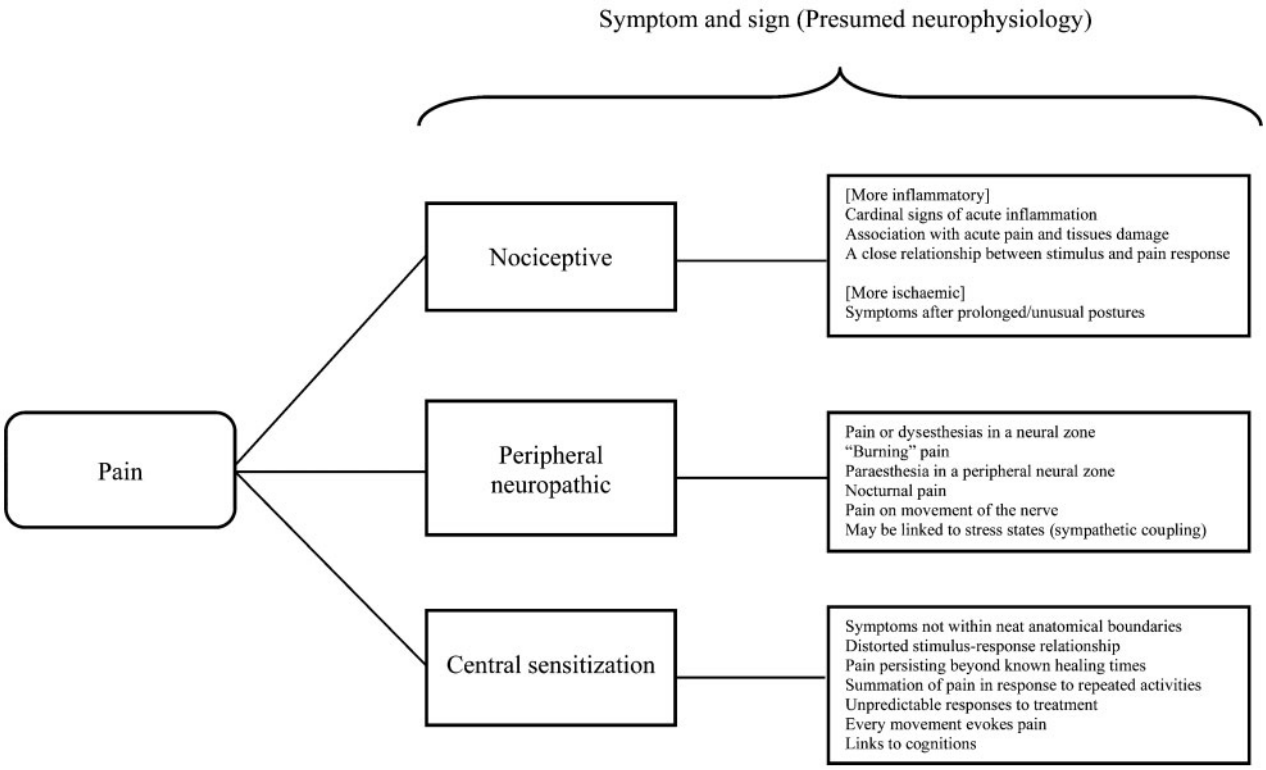

Figure 1 A mechanisms-based classification system for pain, as proposed by Butler.[2]

inviting the selection of specific interventions targeted towards the dominant underlying mechanisms of pain,[7] however this assertion requires further empirical investigation by means of criterion (predictive and prescriptive) validation.

In one approach, three mechanisms-based categorizations relevant to patients with musculoskeletal pain (e.g. low back ± leg pain) have been described including 'nociceptive', 'peripheral neuropathic' and 'central' pain.[2] An outline of the structure of the classification system together with a selected number of category-specific clinical criteria is presented in Fig. 1. In this approach it is proposed that patients' pain may be classified clinically according to the dominant operant mechanisms responsible for its generation and/or maintenance on the basis of known pathophysiology, clinical patterns of symptoms and signs and logic.[2]

The neurophysiological processes underlying each category and a narrative review of the approach have been presented in detail elsewhere.[2,8,9] Briefly, pain that is predominantly 'nociceptive' refers to pain arising predominantly from somatic tissues (e.g. intervertebral disk) in response to inflammation or ischemia associated with injury or loading; 'peripheral neuropathic' pain refers to pain arising predominantly from neural tissue (e.g. nerve root) pathology; and 'central' pain refers to those pain states (e.g. severe, chronic disabling low back pain) generated and/or maintained by an assumed dominance of neurophysiological dysfunction (i.e. nociceptive facilitation and/or dis-inhibition) within the diffuse neural networks of the central nervous system responsible for the transmission and mediation of nociception, leading to persistent pain through 'central sensitization/hyper-excitability'.[1]

In the absence of any criterion or diagnostic gold standards it has been suggested that pain states characterized by a dominance of 'nociceptive', 'peripheral neuropathic' and 'central' mechanisms may be distinguishable from one another clinically, based on the pattern recognition of clusters of symptoms and signs particular to each category.[2,10] Employing a methodology considered suitable for classification system development,[11,12] an expert consensus-derived list of clinical criteria associated with such mechanistic classifications has recently been generated;[13] however, the reliability of clinical judgments and criteria associated with mechanisms-based classifications of pain has not been previously tested. Classification system validation is a multistep process,[14] and proof of reliability is considered a prerequisite of validity,[15] i.e. a classification system cannot be considered valid until its reliability has first been demonstrated, where reliability refers to an assessment procedure with demonstrable reproducibility on repeated administrations of the procedure.[16]

Accepting reliability as a prerequisite of validity, the purpose of this study was to assess the inter- and intra-examiner reliability of clinical judgments associated with: (i) mechanisms-based classifications of pain; and (ii) the identification and interpretation of individual symptoms and signs from a Delphi-derived expert consensus list of clinical criteria associated with mechanisms-based classifications of pain[13] in patients with low back (± leg) pain disorders.

Methods

Setting and participants

This preliminary reliability study was carried out at two Dublin-based university teaching hospitals between April and October 2008. The ethics and medical research committees of each institution reviewed and approved the study protocols.

Participants in the inter-examiner reliability study included a conveniently sampled cohort of 40 patients referred to the Back Pain Screening Clinic of The Adelaide and Meath Hospital (Dublin); the intra-reliability study included a separate cohort of 40 patients referred to the Physiotherapy Department of St Vincent's University Hospital (Dublin). Patients of 18 years of age or older referred with low back (±leg) pain were eligible for inclusion. Exclusion criteria included patients with a history of diabetes, central nervous system injury (e.g. stroke), pregnancy or non-musculoskeletal low back pain. All patients gave signed informed consent prior to their participation in the study. Forty patients were recruited into each study based on sample size estimates for detecting a kappa coefficient of 0.5 with an alpha level of 0.05 and a beta level of 0.2, with a two-tailed hypothesis.[17] Patient demographics for each study are detailed in Table 1.

The examiners consisted of one senior musculoskeletal physical therapist (KS) and one clinical specialist (Back Pain Screening Clinic) (AC) physical therapist, with 11 and 10 years experience respectively as specialists in musculoskeletal physiotherapy. Both examiners possessed 'master of science' level qualifications in physical therapy.

Prior to participation each examiner consulted an assessment manual containing definitions of terms and guidelines for the standardized assessment and interpretation of all subjective and clinical examination criteria. The assessment protocol was developed in part by one examiner (KS) who also delivered a 2-hour assessment protocol training session to the other (AC) in order to clarify and confirm understanding of the assessment procedure.

Procedure

The clinical interview and examination procedures followed a standardized format based on accepted clinical practice.[18] During the clinical interview, patients were encouraged to disclose details of their low back pain history, current symptomology, and its behavior. Patients were also screened for 'red' and 'yellow' flags associated with serious spinal pathology and psychosocial mediators, respectively. The clinical examination included postural, movement, and neurological-based assessments. A number of additional subjective (e.g. spontaneous, paroxysmal, and dysesthetic pain) and clinical examination Delphi-derived criteria (e.g. allodynia, hyperalgesia, hyperpathia, and nerve palpation) were also assessed.

Table 1 Patient demographics

	Inter-observer study (n=40)	Intra-observer study (n=40)
Gender		
Male	20	16
Female	20	24
Age (years)		
Mean	44	47
Range	20–76	24–73
Duration of current episode		
0–6 weeks	6	15
7–12 weeks	4	3
4–6 months	7	2
7–12 months	6	4
>1 year (mean, range in years)	17 (5.5, 1–22)	16 (5.0, 1–18)
Predominant pain location		
Back	24	21
Back/thigh	5	6
Unilateral leg, below knee	7	7
Back + unilateral leg, below knee	2	6
Bilateral leg, below knee	1	0
Back + bilateral leg, below knee	1	0
Work status		
Full-time, full duties	15	7
Full-time, modified duties	3	2
Part-time, full duties	0	3
Part-time, modified duties	0	1
Homemaker	4	2
Retired	3	7
Unemployed	1	2
Off work (weeks): low back pain (mean, range)	12 (20, 1–69)	13 (24, 1–67)
Off work: other	2	2
Registered disabled: low back pain	0	1
Medico-legal case pending		
No	37	38
Yes	3	2

The inter-examiner reliability study employed a simultaneous examiner design, whereby each patient was interviewed and examined by one clinician (AC), whilst the other (KS) observed, in order to ensure that examiners witnessed and interpreted identical patient responses to both the clinical history and examination. This approach was used because of the potential for repeated clinical interviews and tests to lead to variable responses and has been employed in similar reliability studies.[19,20] For the intra-examiner reliability study, each patient was assessed by the same clinician (KS) on two separate occasions (mean number of days between assessments =11, SD =9.0, range 6–56).

After each clinical assessment, examiners were required to:

(i) classify each patient's pain presentation according to a reference standard of expert clinical judgment, i.e. a clinical impression as to the dominant category of mechanisms assumed to underlie each clinical presentation of low back (±leg) pain. Seven response options included 'nociceptive', 'peripheral neuropathic', 'central' or one of four possible 'mixed' pain states derived from a combination of the original three categories;

(ii) complete a 38-item 'clinical criteria checklist' (CCC), consisting of 26 subjective and 12 clinical examination criteria, based on a Delphi survey consensus-derived list of clinical criteria associated with a clinical dominance of 'nociceptive', 'peripheral neuropathic' and 'central' mechanisms of pain (see Tables 2 and 3).[13] Response options for each criterion included 'present', 'absent', or 'don't know'.

Pain classification and completion of the CCC was performed independently by each clinician.

Data analysis

The inter- and intra-examiner reliability of: (i) examiners' mechanisms-based pain classifications; and (ii) each clinical criterion were analyzed using kappa coefficients with two-sided 95% confidence intervals (CIs) and percentage of agreement.[15] The kappa statistic is considered an appropriate method for calculating agreement beyond chance.[21] Percentage of agreement values were also calculated in order to provide a measure of agreement for those criteria where kappa could not be calculated, i.e. where the rating for a criterion was a constant according to either observer or secondary to incomplete cell filling.[22,23] Data were analyzed using the Statistical Package for the Social Sciences (SPSS) for Windows version 14.0 (SPSS Inc., Chicago, IL, USA) and Graphpad (www.graphpad.com/quickcalcs/kappa1.cfm). Interpretations of kappa coefficients were based on distinctions outlined by Landis and Koch.[24] For the purpose of this study 'clinically acceptable' reliability was defined as kappa $\geq$0.61 or in the absence of kappa a percentage agreement of $\geq$80%.[23,25,26]

Results

Inter-examiner kappa and percentage of agreement values for clinical judgments associated with mechanisms-based classifications of low back (±leg) pain were: kappa =0.77 95% CI: 0.57–0.96; and 87.5%. The corresponding intra-examiner values were: kappa =0.96; 95% CI: 0.92–1.00; and 92.5%, suggesting 'substantial' and 'almost perfect' agreement respectively.[24]

Inter- and intra-examiner kappa and percentage of agreement values for each subjective and clinical examination criterion on the 38-item CCC are displayed in Tables 2 and 3. Sixty-eight (26 out of 38) and 95% (36 out of 38) of items on the CCC demonstrated clinically acceptable inter- and intra-examiner reliability respectively.

Discussion

As part of the process towards the development and validation of mechanisms-based classifications of musculoskeletal pain, this study evaluated the reliability of clinical judgments associated with: (i) mechanisms-based classifications of pain; and (ii) a Delphi-derived expert consensus list of clinical criteria associated with mechanisms-based classifications of pain, in a patient population with low back (±leg) pain. For a classification system to be valid it must be reliable, i.e. the clinical judgments and criteria that determine group membership must be reproducible and identifiable in a consistent manner.[27] The inter- and intra-examiner agreement of clinical judgments associated with mechanisms-based classifications of low back (±leg) pain was 'substantial' (kappa =0.77) to 'almost perfect' (kappa =0.96), respectively. The findings from this study provide some preliminary evidence that clinical judgments associated with mechanisms-based classifications of pain may have acceptable clinical reliability.

However the reliability (and validity) of clinical judgments related to mechanisms-based classifications of pain should be interpreted with caution. At present, mechanisms-based classifications undertaken clinically are predicated upon the identification of clusters of symptoms and signs which are in turn assumed to arise from and reflect the dominant pain generating mechanisms.[2,9,28] However, the extent to which clinicians may validly equate clinical phenomenology (i.e. patterns of symptoms and signs) with the underlying pathophysiological pain mechanisms is uncertain.[29,30] This is because clinical pain is invariably the result of multiple pathophysiological mechanisms[31] and because a single pain-related clinical finding (e.g. allodynia) may occur as a result of a number of different mechanisms.[32] In light of these limitations mechanisms-based classifications of pain should at best be viewed as clinically-oriented indirect classifications based on the identification of

Table 2 Inter- and Intra-examiner reliability of the 26 'subjective' items from a 'clinical criteria checklist' associated with mechanisms-based classifications of pain (n=40)

	Inter-examiner			Intra-examiner		
Criterion	**Kappa (95% CI)**	**% agreement**	**Proportion of positive observations (%) examiner 1/examiner 2**	**Kappa (95% CI)**	**% agreement**	**Proportion of positive observations (%) time 1/time 2**
1. Pain of recent onset	1.00 (N/A)	100	13/13	1.00 (N/A)	100	38/38
2. Pain associated with and in proportion to trauma or a pathological process or movement/postural dysfunction	0.55 (0.35–0.75)	90	90/85	1.00 (N/A)	100	85/85
3. History of nerve injury, pathology or mechanical compromise	0.70 (0.58–0.82)	85	50/55	0.90 (0.83–0.97)	95	38/43
4. Pain disproportionate to the nature and extent of injury or pathology	0.63 (0.44–0.82)	92.5	8/15	1.00 (N/A)	100	15/15
5. Usually intermittent and sharp with movement/mechanical provocation; may be a more constant dull ache or throb at rest	0.49 (0.31–0.67)	87.5	93/80	0.77 (0.62–0.93)	95	88/88
6. More constant/unremitting pain	0.63 (0.44–0.82)	92.5	8/15	0.68 (0.51–0.85)	92.5	15/13
7. Pain variously described as burning, shooting, sharp, or electric-shock-like	0.33 (0.20–0.47)	70	25/48	0.70 (0.59–0.81)	85	50/45
8. Pain localized to the area of injury/dysfunction (with/without some somatic referral)	0.40 (0.26–0.53)	72.5	58/80	0.76 (0.63–0.89)	92.5	78/85
9. Pain referred in a dermatomal or cutaneous distribution	0.65 (0.53–0.77)	82.5	48/40	0.79 (0.68–0.91)	92.5	25/23
10. Widespread, non-anatomical distribution of pain	–	95	3/3	0.84 (0.69–1.00)	97.5	8/10
11. Clear, proportionate mechanical/anatomical nature to aggravating and easing factors	0.47 (0.16–0.79)	95	95/95	0.90 (0.79–1.00)	97.5	85/88
12. Mechanical pattern to aggravating and easing factors involving activities/postures associated with movement, loading or compression of neural tissue	0.95 (0.90–1.00)	97.5	43/45	0.89 (0.81–0.96)	97.5	30/35
13. Disproportionate, non-mechanical, unpredictable pattern of pain provocation in response to multiple/non-specific aggravating/easing factors	–	97.5	0/3	1.00 (N/A)	100	10/10
14. Reports of spontaneous (i.e. stimulus-independent) pain and/or paroxysmal pain (i.e. sudden recurrences and intensification of pain)	0.80 (0.71–0.89)	90	50/45	0.88 (0.79–0.96)	95	30/25
15. Pain in association with other dysesthesias (e.g. crawling, electrical, heaviness, coldness, burning)	0.88 (0.80–0.96)	95	30/30	0.72 (0.54–0.91)	95	10/10
16. Pain of high severity and irritability (i.e. easily provoked, taking a long time to settle)	0.36 (0.17–0.55)	82.5	18/15	1.00 (N/A)	100	5/5
17. Pain in association with other symptoms of inflammation (i.e. swelling, redness, heat) (inflammatory nociceptive)	–	97.5	3/0	–	97.5	3/0
18. Pain in association with other neurological symptoms (e.g. pins and needles, numbness, weakness)	0.85 (0.76–0.93)	92.5	45/38	0.88 (0.80–0.96)	95	33/28
19. Night pain/disturbed sleep	0.84 (0.75–0.93)	92.5	68/60	0.71 (0.58–0.85)	90	25/20
20. Responsive to simple analgesia/NSAIDs*	0.49 (0.18–0.81)	80	80/70	0.43 (0.26–0.60)	80	75/80
21. Less responsive to simple analgesia/NSAIDs and/or more responsive to anti-epileptic (e.g. Neurontin, Lyrica)/anti-depression (e.g. Amitriptyline) medication	0.75 (0.42–1.00)	95	8/13	0.65 (0.52–0.78)	87.5	13/13
22. Rapidly resolving or resolving in accordance with expected tissue healing/pathology recovery times	0.55 (0.40–0.70)	82.5	28/25	0.80 (0.71–0.89)	90	45/50
23. Pain persisting beyond expected tissue healing/pathology recovery times	0.60 (0.47–0.72)	80	48/43	0.89 (0.81–0.96)	95	33/33
24. History of failed interventions (medical/surgical/therapeutic)	0.90 (0.79–1.00)	97.5	15/13	0.88 (0.79–0.96)	95	28/28
25. Strong association with maladaptive psychosocial factors (e.g. negative emotions, poor self-efficacy, maladaptive beliefs and pain behaviours, altered family/work/social life, medical conflict)	0.42 (0.22–0.61)	85	18/13	0.73 (0.58–0.88)	92.5	18/15
26. Pain in association with high levels of functional disability	0.16 (−0.02–0.34)	82.5	5/18	0.88 (0.79–0.96)	95	30/25

Note: CI, confidence interval.
*Non-steroidal anti-inflammatory drugs.

Table 3 Inter- and Intra-examiner reliability of the 12 'clinical examination' items from a 'clinical criteria checklist' associated with mechanisms-based classifications of pain (n=40)

	Inter-examiner			Intra-examiner		
Criterion	**Kappa (95% CI)**	**% agreement**	**Proportion of positive observations (%) examiner 1/examiner 2**	**Kappa (95% CI)**	**% agreement**	**Proportion of positive observations (%) time 1/time 2**
27. Antalgic (i.e. pain relieving) postures/movement patterns	0.66 (0.48–0.81)	90	20/15	0.72 (0.54–0.91)	95	13/8
28. Clear, consistent and proportionate mechanical/anatomical pattern of pain reproduction on movement/mechanical testing of target tissues	0.29 (0.03–0.54)	90	90/95	0.77 (0.62–0.93)	95	88/88
29. Pain/symptom provocation with mechanical/movement tests that move/load/compress neural tissue (e.g. active/passive, neurodynamic, i.e. straight leg raise)	0.75 (0.65–0.85)	90	48/50	0.95 (0.90–1.00)	97.5	38/40
30. Disproportionate, inconsistent, non-mechanical/non-anatomical pattern of pain provocation in response to movement/mechanical testing	–	100	0/0	0.64 (0.40–0.88)	95	8/8
31. Positive neurological findings (including altered reflexes, sensation and muscle power in a dermatomal/myotomal or cutaneous nerve distribution)	0.94 (0.88–1.00)	97.5	28/30	0.74 (0.62–0.86)	90	30/20
32. Localized pain on palpation	0.78 (0.68–0.88)	90	65/65	1.00 (N/A)	100	75/75
33. Diffuse/non-anatomic areas of pain/tenderness on palpation	–	95	0/5	1.00 (N/A)	100	8/8
34. Positive findings of allodynia within the distribution of pain	1.00 (N/A)	100	3/3	0.66 (0.34–0.98)	97.5	5/3
35. Positive findings of hyperalgesia (primary, secondary) within the distribution of pain	0.86 (0.76–0.95)	95	20/25	0.69 (0.54–0.83)	90	20/20
36. Positive findings of hyperpathia within the distribution of pain	0.94 (0.89–1.00)	97.5	33/30	0.53 (0.37–0.70)	85	23/18
37. Pain/symptom provocation on palpation of relevant neural tissues	0.82 (0.71–0.92)	92.5	30/28	0.83 (0.71–0.95)	95	15/20
38. Positive identification of various psychosocial factors (e.g. catastrophization, fear-avoidance behaviour, distress)	0.63 (0.44–0.82)	92.5	15/8	0.83 (0.71–0.95)	95	20/15

Note: CI, confidence interval.

symptoms and signs presumed to reflect the underlying dominant pain generating mechanisms rather than direct classifications based on neurophysiology.

As an alternative to clinical judgment, a number of generic[33] and back-pain specific[34] screening instruments for use in clinical practice have been developed in order to help clinicians identify a dominance of (peripheral) neuropathic pain. The propensity of these tools to dichotomize pain as arising from a dominance of either 'nociceptive' or 'peripheral neuropathic' mechanisms limits their ability to screen for the presence of other potentially dominant pain states such as central hyper-excitability/sensitization[1] for which validated screening instruments do not currently exist.

Results from the inter-examiner reliability study revealed clinically acceptable reliability for 58% (n=15) of the 26 subjective items on the CCC as compared to 96% (n=25) in the intra-observer study. These figures contrast with those associated with the 12 'clinical examination' criteria where 92% (n=11) of criteria demonstrated equivalent levels of inter- and intra-examiner agreement. This finding could suggest that clinicians may interpret patients' responses to clinical tests more reliably than their responses to clinical interview questions.

This preliminary reliability study found varying ('slight' to 'almost perfect') levels of inter-examiner agreement associated with the clinicians' interpretations of patients responses to clinical interview questions designed to elicit those symptoms associated with 'nociceptive', 'peripheral neuropathic', and 'central' pain. For example, the inter-examiner reliability of neuropathic pain descriptors evaluated in the present study (criterion 7) was 'fair' (kappa =0.33). In an inter-tester reliability study of the clinical questions and tests associated with diagnostic triage of non-specific low back pain involving 301 patients and 27 pairs of independent examiners, McCarthy *et al.*[35] found moderate levels (kappa =0.46–0.58) of inter-tester reliability associated with the identification of qualitative pain descriptors (dull, sharp, deep, superficial, throbs) deemed suggestive of nerve root pathology (i.e. peripheral neuropathic pain). According to our criteria (kappa $\geqslant$0.61), these findings suggest that the identification of neuropathic descriptors by clinicians may be clinically unreliable. Should this finding be replicated it may suggest that screening instruments rather than individual clinical judgments should be used to screen for the presence of (peripheral) neuropathic pain.

The inter-examiner reliability of clinical judgments associated with the geography of patients' pain (e.g. criterion 8: localized; criterion 9: dermatomal distribution) was 'fair' to 'substantial' (kappa =0.40–0.65). This finding is comparable to that reported by McCarthy *et al.*[35] whose examiners were required to discern between a dominance of leg (kappa =0.57) versus back pain (kappa =0.46). Together, these findings suggest that clinical judgments regarding leg/dermatomal pain may be more reliable than judgments related to the localized nature of low back pain. However, lower reliability levels associated with the identification of radicular distributions of pain have also been reported. In a study of the inter-observer reliability of symptoms and signs associated with nerve root pathology involving 91 patients and two pairs of independent examiners, Vroomen *et al.*[36] found 'fair' agreement (kappa = 0.24) between clinicians ability to identify a 'typical radiation pattern of pain'. Taken together and by the criteria employed in this study, these findings suggest that overall the reliability of clinical judgments related to the geography of patients' pain may be clinically unacceptable.

International guidelines recommend the assessment of psychosocial risk factors (e.g. maladaptive beliefs and behaviours; family, work and social issues) as prognostic indicators for poor recovery from episodes of low back pain.[37,38] The inter-examiner agreement of a general criterion (criterion 25) related to the identification of maladaptive psychosocial issues in this study was moderate (kappa =0.42), a finding consistent with the 'fair to moderate' (kappa =0.29–0.58) levels of inter-tester agreement associated with various individual psychosocial 'yellow flags' found by others.[35] Together these data suggest that clinicians may not be able to identify psychosocial risk factors from the clinical interview with a clinically acceptable degree of reliability. Considering their importance as prognostic indicators,[39] upon which important clinical decisions may be based, the reasons for this clinically insufficient level of reliability warrants further exploration. As with neuropathic pain, existing screening instruments may assist clinicians to identify psychosocial risk factors more reliably.

A recent systematic review of the reliability of clinical examination procedures in non-specific low back pain found generally low levels of reliability associated with palpation and symptom response-based examination procedures.[40] However, where clinical criteria are considered to have important clinical value then efforts may be required to: (i) standardize and operationalize the criterion in a way that improves its reliability whilst maintaining its diagnostic/clinical value; and (ii) improve training in a way that allows clinicians to assess the criterion more reliably.[41]

This pilot study found varying ('fair' to 'almost perfect') levels of inter-examiner agreement associated with clinicians' interpretations of patients responses to clinical tests associated with 'nociceptive', 'peripheral neuropathic' and 'central' pain. For

example, 'antalgic (i.e. pain relieving) postures/movement patterns' (e.g. Gower's sign[42] or holding a limb in a flexed position in order to unload sensitive neural tissues[43] have been suggested as potential clinical indicators associated with a dominance of nociceptive and peripheral neuropathic pain.[13] Comparative data concerning the reliability of clinical observations related to antalgic postures and/or movement patterns is limited and equivocal. Whereas the preliminary finding from this pilot study provides some evidence of 'substantial' inter-examiner (kappa = 0.66) and intra-examiner (kappa = 0.72) reliability a limited body of contradictory data suggests observations of antalgic 'postures', such as a laterally shifted lumbar spine (kappa =0.00–0.53),[40] or in patients with suspected lumbar nerve root involvement, a flexed knee whilst standing (kappa =0.38),[36] may lack reliability as observable features associated with low back (nociceptive) or nerve root (peripheral neuropathic) pain. Further studies are required in order to clarify the uncertainty.

The inter- and intra-examiner reliability of the clinicians interpretations of pain/symptom provocation associated with neural tissue provocation tests (e.g. straight leg raise (SLR), slump) was 'substantial' (kappa =0.75) and 'almost perfect' (kappa =0.95) respectively. These findings are broadly consistent with those reported by McCarthy *et al.*[35] (kappa = 0.67). Additional evidence supporting the reliability of the SLR has been reported. Bertilson *et al.*[44] evaluated the reliability of the 30 clinical tests for the lumbar spine including the SLR in an inter-examiner study involving one pair of independent examiners and 50 patients and found an 'almost perfect' level of agreement (kappa =0.92). Similarly, van Dillen *et al.*[20] evaluated the reliability of the SLR as part of a battery of physical examination items involving 95 patients with low back pain using a simultaneous examiner design and found 'almost perfect' agreement (kappa =0.93). Overall, the SLR appears to have clinically acceptable reliability as a pain provocation test.

Limitations

The findings from this reliability study should be considered preliminary and interpreted in light of a number of methodological limitations. The simultaneous examiner design may have introduced bias towards inflated levels of inter-examiner agreement since it does not take account of the potential variability's in assessments and patient–therapist interactions that would otherwise occur in an independent-examiner design and which might be expected to produce lower levels of agreement. Therefore, the reliability of clinical judgments linked to mechanisms-based classifications and associated clinical criteria should be tested using a more robust design employing a larger patient sample and multiple independent examiners.[40] Also, the levels of agreement associated with the intra-examiner study could be inflated secondary to recall bias and the fact that one of the raters helped develop the system and thus had intimate knowledge of it. Also, the examiners in this study had over 10 years experience in musculoskeletal-based physical therapy assessments. As such the results of this study may not generalize to clinicians less acquainted with mechanisms-based approaches or those with less experience. This study has endeavoured to make some comparisons with reliability data reported in the relevant literature; however, direct comparisons may be inappropriate considering differences in experimental design, statistical analyses, assessment scales, patient populations, the prevalence of criteria/disease states and bias between raters.[17,27]

Conclusion

This study provides some preliminary evidence supporting the reliability of clinical judgments associated with mechanisms-based classifications of pain in patients with low back (±leg) pain disorders. Mixed levels of agreement were demonstrated for the 38-item 'CCC' upon which such classifications may be based. Further empirical studies evaluating the reliability of mechanisms-based classifications of pain using larger patient samples and multiple independent examiners are required to justify this approach in clinical practice.

Acknowledgements

This research was funded by the Health Research Board (of Ireland) (grant no. CTPF-06-17).

References

1 Lidbeck J. Central hyperexcitability in chronic musculoskeletal pain: a conceptual breakthrough with multiple clinical implications. Pain Res Manag. 2002;7:81–92.
2 Butler DS. The sensitive nervous system. Adelaide: NOI Group Publications; 2000.
3 Woolf CJ, Bennett GJ, Doherty M, Dubner R, Kidd B, Koltzenburg M, *et al.* Towards a mechanism-based classification of pain? Pain. 1998;77:227–9.
4 Dallel R, Voisin D. Towards a pain treatment based on the identification of the pain generating mechanisms? Eur Neurol. 2001;45:126–32.
5 Gifford L. Pain, the tissues and the nervous system: a conceptual model. Physiotherapy. 1998;84:27–36.
6 Haldeman S. North American Spine Society: failure of the pathology model to predict back pain. Spine. 1990;15:718–24.
7 Baron R. Mechanisms of disease: neuropathic pain-a clinical perspective. Nat Clin Pract Neurol. 2006;2:95–106.
8 McMahon SB, Koltzenburg M. Wall and Melzack's textbook of pain. 5th ed. Philadelphia (PA): Elsevier/Churchill Livingstone; 2006.
9 Smart KM, O'Connell NE, Doody C. Towards a mechanisms-based classification of pain in musculoskeletal physiotherapy? Phys Ther Rev. 2008;13:1–10.
10 Finnerup NB, Jensen TS. Mechanisms of disease: mechanism-based classification of neuropathic pain – a critical analysis. Nat Clin Pract Neurol. 2006;2:107–15.
11 Ford J, Story I, O'Sullivan P, McMeeken J. Classification systems for low back pain: a review of the methodology for development and validation. Phys Ther Rev. 2007;12:33–42.
12 Katz JN, Stock SR, Evanoff BA, Rempel D, Moore JS, Franzblau A, *et al.* Classification criteria and severity assessment in work-associated upper extremity disorders: methods matter. Am J Ind Med. 2000;38:369–72.

13 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. 2010;15:80–7.
14 Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.
15 Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. London: Pearson Prentice Hall; 2008.
16 Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport. 2000;1:54–61.
17 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85:257–268.
18 Petty NJ, Moore AP. Neuromusculoskeletal examination and assessment. A handbook for therapists. Edinburgh: Churchill Livingstone; 2001.
19 Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C, *et al*. Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. Aust J Physiother. 2004;50:85–94.
20 van Dillen LR, Sahrman SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK, *et al*. Reliability of physical examination items used for classification of patients with low back pain. Phys Ther. 1998;78:979–88.
21 Daly LE, Bourke GJ. Interpretation and uses of medical statistics. 5th ed. Oxford: Blackwell Science Ltd; 2000.
22 Clarkson Keller VE, Keck JF. An instrument for observational assessment of nausea in young children. Pediatr Nurs. 2006;32:420–6.
23 Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JM, Gröken LN. Test-retest reliability of the Isenhargen work systems functional capacity evaluation in patients with chronic low back pain. J Occup Rehabil. 2003;13:207–18.
24 Landis R, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
25 Nomden JG, Slagers AJ, Bergman GJD, Winters JC, Kropmans TJB, Dijkstra PU. Interobserever reliability of physical examination of the shoulder girdle. Man Ther. 2009;14:152–9.
26 Hunt DG, Zuberbier OA, Kozlowski AL, Robinson J, Berkowitz J, Schultz JZ, *et al*. Reliability of the lumbar flexion, lumbar extension, and passive straight leg raise test in normal populations embedded within a complete physical examination. Spine. 2001;26:2714–8.
27 Fritz JM, Brennan GP, Clifford SN, Hunter SJ, Thackeray A. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. Spine. 2006;31:77–82.
28 Woolf CJ, Max MB. Mechanisms-based pain diagnosis. Anesthesiol. 2001;95:241–9.
29 Hansson P. Difficulties in stratifying neuropathic pain by mechanisms. Eur J Pain. 2003;7:353–7.
30 Dworkin RH. An overview of neuropathic pain: syndromes, symptoms, signs and several mechanisms. Clin J Pain. 2002;18:343–9.
31 Max MB. Is mechanism-based pain treatment attainable? Clinical trial issues. J Pain. 2000;1(Suppl 1):2–9.
32 Finnerup NB, Jensen TS. Mechanisms of disease. mechanism-based classification of neuropathic pain – a critical analysis. Nat Clin Pract Neurol. 2006;2:107–15.
33 Bennett MI, Attal N, Backonja MM, Baron R, Bouhassira D, Freynhagen R, *et al*. Using screening tools to identify neuropathic pain. Pain. 2007;127:199–203.
34 Freynhagen R, Baron R, Gockel U, Tölle TR. Pain detect. A new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006;22:1911–20.
35 McCarthy CJ, Gittins M, Roberts C, Oldham JA. The reliability of the clinical tests and questions recommended in the international guidelines for low back pain. Spine. 2007;32:921–6.
36 Vroomen PCAJ, de Krom MCTFM, Knottnerus JA. Consistency of history taking and physical examination in patients with suspected lumbar nerve root involvement. Spine. 2000;25:91–7.
37 COST B13 Working Group. European guidelines for the management of acute nonspecific low back pain in primary care. 2005. Available from: http://www.backpainineurope.org/web/html/wg1_results.html.
38 COST B13 Working Group. European guidelines for the management of chronic nonspecific low back pain. 2004. Available from: http://www.backpainineurope.org/web/html/wg2_results.html.
39 Pincus T, Burton KA, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. Spine. 2002;27:E109–20.
40 May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain. A systematic review. Aust J Physiother. 2006;52:91–102.
41 Lurie JD. What diagnostic tests are useful for low back pain? Best Pract Res Clin Rheumatol. 2005;19:557–75.
42 Cook C, Brismée JM, Sizer PS. Subjective and objective descriptors of clinical lumbar spine instability: a Delphi study. Man Ther. 2006;11:11–21.
43 Nee RJ, Butler D. Management of peripheral neuropathic pain: integrating neurobiology, neurodynamics, and clinical evidence. Phys Ther Sport. 2006;7:36–49.
44 Bertilson BC, Bring J, Sjöblom A, Sundell K, Strender LE. Inter-examiner reliability in the assessment of low back pain (LBP) using the Kirkaldy-Willis classification (KWC). Eur Spine J. 2006;15:1695–703.