J Rehabil Med 2006; 38: 263–267

Taylor & Francis
Taylor & Francis Group

DEVELOPMENT OF A CLINICAL EXAMINATION IN NON-SPECIFIC LOW BACK PAIN: A DELPHI TECHNIQUE

Christopher J. McCarthy, PhD[1], Alison Rushton, EdD[2], Vicky Billis, MSc[1], Frances Arnall, MPhil[1] and Jacqueline A. Oldham, PhD[1]

From the [1]Centre for Rehabilitation Science, University of Manchester, Manchester Royal Infirmary, Manchester and [2]School of Health Sciences, University of Birmingham, Birmingham, UK

Objective: **To establish the discriminatory items of the clinical examination of non-specific low back pain, important to physiotherapists.**
Design: **A focus group and Delphi technique with UK physiotherapists.**
Subjects: **A purposive sample of 30 physiotherapists attended a focus group and completed 3 rounds of Delphi questionnaires.**
Methods: **Data were analysed using mixed qualitative and quantitative approaches. A frequency content analysis identified commonly identified tests and questions, whilst the Delphi consensus technique assumed consensus had been reached with greater than 80% agreement on item inclusion or exclusion.**
Results: **The focus group established the structure of the clinical examination with 15 domains of questioning or physical testing. Three rounds of Delphi questionnaires established the important items of the clinical examination. The list of tests and questions included items evaluating both the psychosocial and biomedical status of the patient as well as questions screening for red flags.**
Conclusion: **This is the first work to establish discriminatory tests in the clinical examination of non-specific low back pain, important to physiotherapists. The clinical examination will subsequently be evaluated for item validity and data will undergo cluster analysis. The items of this clinical examination may provide evidence for the existence of homogenous subgroups within the heterogeneous non-specific low back pain diagnosis.**

Key words: back pain, classification, Delphi technique.

J Rehabil Med 2006; 38: 263–267

Correspondence address: Christopher McCarthy, The Centre for Rehabilitation Science, University of Manchester, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL, UK. E-mail: christopher.j.mccarthy@man.ac.uk

Submitted July 20, 2005; accepted February 14, 2006

INTRODUCTION

Patients with non-specific low back pain (NSLBP) are commonly referred for assessment and treatment by a physiotherapist. Physiotherapists use a range of conservative treatments in their management of back pain, ranging from massage and mobilization to back schools and multidisciplinary group programmes (1). Unfortunately, evidence supporting the use of the majority of physiotherapeutic treatments is sometimes contradictory (2). The authors of the most comprehensive review of the evidence for conservative treatment yet performed (2), agreed with previous authors (3, 4) on a major limitation in the literature in this field, namely the heterogeneity of the condition.

It has long been recognized that the group of patients referred for conservative treatment with the diagnosis of NSLBP form an extremely heterogeneous group and that prognosis and optimal treatment methods vary immensely within this group (4). It is considered that this heterogeneous group consists of several smaller homogenous subsets, with each subset being more likely to respond to a type of treatment unique to that subset (4). Thus, with the recognition that particular conservative treatments may be more efficacious with certain subsets of patients, than for the whole heterogeneous group of LBP sufferers, there has been a strong international recommendation to establish a method of classification that will distinguish one subset from another (2).

Physiotherapists undertake an informal process of "diagnostic" classification during their clinical examination, involving the use of signs and symptoms and pattern recognition models of clinical reasoning (5), whilst several authors have developed classification systems based on sub-classifying the patient with NSLBP according to clinical history, pain presentation, movement dysfunction and psychosocial influences (6). These classifications suffered from being designed based on the judgement of small numbers of clinicians rather than on a representative professional consensus. Recently, work has been undertaken that incorporates elements of the McKenzie examination with other clinical tests, demonstrating some evidence of criterion-related validity; however, more work is needed (7–9).

The other commonly adopted method of classification development is to use a statistical cluster analysis approach (10, 11). Here the investigator typically conducts extensive testing, obtains comprehensive data from subjects and then enters the data into a statistical model that enables the identification of clusters of signs and symptoms. These clusters of features have the advantage of being developed using a rigorous methodology, but may not necessarily portray clinically recognizable syndromes. This is the case particularly if the tests and questions on which the statistical analysis is based do

© 2006 Taylor & Francis. *ISSN 1650-1977*
DOI: 10.1080/16501970600632768

not possess adequate measurement validity. Thus, whilst the use of a statistical approach has been shown to produce methodologically rigorous classification systems (12), the clusters produced are only as valid as the particular examination questions and tests used in the analysis.

One approach to addressing the limitations in validity of some statistically derived clusters is to develop a consensus on the important discriminatory clinical tests and questions that should be entered into the statistical cluster model. Thus, by using features of the examination derived by a professional consensus, the output from a statistical cluster analysis should consist of subgroups that have strong content and construct validity and are recognizable to the majority of clinicians. This study aimed to establish consensus, amongst UK Chartered Physiotherapists, on the items of a lumbar clinical examination that will subsequently be subject to statistical analysis for their ability to sub-classify NSLBP.

METHODS

The present study involved 2 stages; a focus group followed by 3 rounds of Delphi consensus technique questionnaires. The study received a favourable ethical decision from the South-East multi-centre research ethics committee (MREC).

Stage 1: focus group

Focus groups are a form of group interview that capitalizes on communication between research participants and are commonly used to explore participant opinion and knowledge (13).

Participants

A purposive sampling approach was taken in the selection of participants for the focus group, as in order to best represent professional consensus it was considered important to include the views of all specialties of physiotherapist, thus ensuring theoretical representativeness. Each Clinical Interest Group (CIG) of the Chartered Society of Physiotherapy was contacted and invited to send a representative to the focus group. CIG represent a distinct field of physiotherapy clinical practice that may be specific to a client group, a clinical area or a specific treatment approach or modality.

Of the 37 contacted, 29 sent a representative to the focus group. The remaining CIGs were contacted and the suitability of sending a representative was discussed. The CIGs who felt it was unnecessary to send representatives felt this was due to their clinical work not bringing them into contact with patients with NSLBP. Participants worked in a mix of work settings, including private practice, primary care, secondary care and lecturers and research staff. In addition, user involvement was achieved through a patient representative, contacted via the BackCare charity patient network, participating in the focus group. It was decided to limit the size of the sample to 30 participants, as greater numbers may lead to difficulties co-ordinating discussions in the focus group.

Procedure

The focus group participants ($n=30$) gave their written consent to participate in the focus group and agreed that the group's consensus could be discussed beyond the group members. The focus group was lead by a facilitator who was not on the research team and had no vested interest in the study. Additionally, 2 observers took notes during the discussions regarding participant interaction and the group dynamics. After a session where participants and the research team introduced themselves, ground rules for the discussions were set. The use of ground rules encourages open debate and prevents animosity and aggression influencing group discussions. The group decided that it would be appropriate to discuss which areas of the clinical examination were important in discriminating different "types" of NSLBP. The initial discussion involved participants debating which areas of the history and physical examination should be included in the list.

Having discussed the areas, participants were asked to vote for the suggested areas as being "useful", "important" or "very important" in the discrimination of different types of NSLBP. The areas of the examination voted as important or very important in the discrimination of NSLBP can be seen in Tables I and II. Agreement was obtained on the decision to exclude areas voted only as "useful" from the examination list.

Stage 2: Delphi Questionnaire

The Delphi technique is used for canvassing opinion and structured decision-making and consists of consecutive rounds of questionnaires designed to achieve increasing consensus of opinion (14). The technique has been used widely in the medical and nursing fields and has been suggested for use when seeking national opinion and developing priority issues (14). This iterative process allowed participants to suggest items to be included in the examination list and then to view the common items suggested by the group. Subsequent rounds allowed participants to vote for the continued inclusion or exclusion of items in the knowledge of the group's opinion.

Participants

Focus group participants were asked to nominate a further 3 colleagues from their CIG to join the Delphi process, to potentially expand the sample size to 120, however only 4 nominations were received whilst 4 of the original participants withdrew due to time commitments. Thus, the sample size remained at 30.

Procedure

Round 1. In the first round of questionnaires, participants ($n=30$) were provided feedback from the focus group and were asked to suggest "important" discriminatory questions and tests for each area of the examination. Participants were asked to consider any evidence of measurement validity for the tests they were suggesting. Suggestions were collated and a content analysis (15) of the free text responses was conducted independently by 2 of the research team. Commonly suggested questions and tests were coded and a quantitative frequency analysis was conducted. Agreement between the raters was excellent (intra-class correlation coefficient for frequency count, $ICC_{2,1}=0.99$, 95% confidence interval (CI) 0.97–0.99); however, any disagreements in coding were mediated by a third rater.

Round 2. In the second round of questionnaires, the participants were required to vote on whether the item should be included or excluded from the examination list. Participants were asked to agree or disagree on an item's inclusion using a 5-point Likert-scaled question, where 1 was "completely disagree" and 5 "completely agree". There is no definitive level of agreement, with some authors accepting consensus to have been reached with 50% agreement, whilst others have chosen to use 80% (16). Prior to data analysis, it was decided that an item would be included in the list if approximately 75% of participants agreed that it should be included. A margin of variability of +/− 5% was agreed so that an item would be excluded if less than 70% agreed on inclusion and included if 80% agreed on inclusion. Items falling between these margins would be sent round again for reconsideration.

Round 3. In the third round, participants were presented with descriptive statistics (median scores) and percentage agreement for inclusion for each item and were asked to rate the items again in light of this knowledge. Following this round all items had either been included or excluded and consensus on the features of the examination list had been achieved. The results of the third round were fed back to the participants and comments invited. A summary of the items included after the completion of the final round can be seen in Table I and the items that were excluded in Table II.

Table I. *The areas of the clinical examination and a summary of the items generated after the third round of the Delphi*

History	Physical
Pain	Observation
• Mannequin/Body chart use	• Postural assessment
• Area of pain/Description	• General observations
Symptoms other than pain	• Gait assessment
• Weakness/Altered sensation	Physical function
Function	• Demonstration of the functional movement that reproduces symptoms
• Effects on work	Active movements
• Effects on hobbies	• Range and symptom production for Lumbar spine and hips
• Effects on activities of daily living	• Assessment of centralization of pain* (7)
Present history	Passive movements
• Diurnal patterns	• Central pressure at each lumbar level
• Status changing?/Investigations/Treatment	• Pressure on the sacrum at base and apex
Previous history	Muscle assessment
• Previous treatment	• Ability to contract transversus abdominis†
• Expectations of treatment	• Ability to contract multifidus† (24)
Medical considerations	• Control of pelvis whilst moving legs
• "Red flags"	Neurological examination
Psychosocial considerations	• Myotomes
• Attitudes/Beliefs	• Dermatomes
• Behaviours	• Reflexes
• Compensation	• Sensation
• Diagnosis and treatment	• Straight leg raise test. Femoral nerve test‡ (25)
• Emotional issues	Pain behaviour signs
• Family issues	• Behavioural conflict/Exaggeration
• Work issues	

*Progressive retreat of referred pain towards the spinal midline in response to repeated movement testing.
†Isometric contraction of the deep trunk musculature.
‡Neurogenic pain provocation tests.

RESULTS

Attrition rates and incidences of missing data were very low. The first round questionnaire was completed by 29 (97%) of participants, the second round 29 (97%) and the third round 28 (93%). The resultant patient examination list is structured into 2 sections. The patient is initially interviewed with attention being paid to symptomatology, present history, previous history, medical screening questions and an assessment of the psychosocial influences on presentation. The physical examination subsequently assesses the posture and gait of the patient, the active and passive movements of the lumbar spine, the co-ordination of muscle activity around the lumbo-pelvis and neurological status. An assessment of the patient's behavioural responses during the examination is also made.

DISCUSSION

The clinical tests and questions that were considered as important discriminatory features by this group of physiotherapists will appear familiar to the majority of practitioners who routinely examine patients with NSLBP. The process led to the development of an examination that was structured into 2 parts; the history and the physical examination.

The format of the examination and the items included in it, are seen in a number of textbooks used at both undergraduate and postgraduate levels (17). The clinical features can be considered as typical examination features that should be recognizable to most practitioners in the field. During the Delphi process, a number of items, suggested in the first round of questionnaires, were lost from the final list. Whilst, a little surprisingly, aggravating and easing factors for pain were not included in the examination, no items were voted out of the history section of the examination.

A number of tests were voted out of the physical examination. Tests such as the repeated movements (18) and combined movement tests (19), whilst being taught on some undergraduate courses, are typically skills developed at a postgraduate level and thus may not have been considered as important by participants who had no experience of their use. The same could be said for the examination of trigger points and the sacroiliac joint, although the precise reasons for why items were included or excluded from the list is impossible to establish. It is clear that the participants in this study included in the examination list a wide range of tests and questions, likely to represent elements of their training and personal experience of examining patients with NSLBP in their own clinical settings.

The items in the list appear to allow the assessment of biomedical, psychological and social influences on the patient's NSLBP and whilst producing a lengthy examination do not require specialist, postgraduate training to perform. Whilst participants were asked to consider the validity of the clinical tests, a number of the tests included in the list have been shown to possess poor inter-rater reliability and to have little discriminatory or prognostic ability. Participants were aware that

Table II. *Items removed from the examination list by the end of the third round of the Delphi*

History	Physical
Pain • No items voted off the list (aggravating and easing factors not included in the initial rounds of the Delphi process) Symptoms other than pain • No items voted off the list Function • No items voted off the list Present history • No items voted off the list Previous history • No items voted off the list Medical considerations • No items voted off the list Psychosocial considerations • No items voted off the list	Observation • No items voted off the list Physical function • Transfers • Balance tests • Demonstrations of activity of daily living Active movements • Repeated movements (26)* • Combined movements† (27) • Movements of the thoracic spine • Specific sacroiliac tests • Passive physiological intervertebral movements‡ (28) Passive movements • Specific sacroiliac tests • Palpation of nerves (25) Muscle assessment • Trigger point assessment§ (29) • Assessment of muscle tone Neurological examination • Tone • Co-ordination • Autonomic assessment Pain behaviour signs • Waddell's signs¶ (30)

*Repeated flexion, extension, side bending or rotation and its effect on pain.
†Expanded examination of active movements to explore three-dimensional combinations.
‡Passive examination of individual spinal motion segments to movement abnormality.
§Assessment for tender points with muscle that refer pain on palpation.
¶Series of observations during the examination, identifying elements of "pain behaviour".

in the next phase of this study the inter-tester reliability of all the items included in the examination list would be tested and this may have lead the participants to be more inclusive in the inclusion of potentially unreliable testing procedures.

Alternatively, it is possible that some participants were not aware of the strength of evidence for the measurement validity of some of the suggested tests, as some tests, with evidence for their validity, such as some specific sacroiliac stress tests (20) and centralization of leg pain with repeated movement testing (21) were not included. Consequently, the items of the examination list generated by this method differs from the most recently derived examination, described by Petersen et al. (9). These authors developed a classification system based on the reported reliability and validity evidence for individual clinical tests and have demonstrated the system's reliability on patients. It is possible that a compilation of features of the current work and this work may lead to an optimal classification system.

Previously proposed classification systems have been developed using a consensus-based approach. Binkley et al. (22) described a professional consensus on the classification of LBP, derived from a Delphi procedure. However, the system aimed to classify LBP based on the diagnostic labels suggested by a group of physiotherapy experts and made no attempt to validate the system. Only one system was developed from a consensus of a multidisciplinary committee, informed by a systematic review of the evidence. The Quebec Task Force (QTF) classification (23) incorporated multiple dimensions of the patient's illness presentation and whilst making only rudimentary reference to psychological and social influences did provide a system that offered a basic framework for the classification of LBP. However, the QTF classification did not aim to sub-classify NSLBP and has been superseded by more recent developments in the classification of LBP, such as the triaging process recommended in many clinical guidelines. Thus, the present study is the first to establish a consensus on the items of a clinical examination that should be tested on patients and subsequently analysed using cluster analysis modelling.

In conclusion, the present study establishes the items of the clinical examination that are rated as important discriminatory items in the clinical examination of NSLBP. The items are simple questions and physical tests familiar to physiotherapists. They are also likely to be familiar to other clinicians who conduct detailed clinical examinations of patients with NSLBP. The items are predominantly biomedical assessments, but psychological, behavioural and social influences on the patient's presentation are included.

The study must be viewed in light of its limitations. Firstly, development by consensus may have led to some apparent anomalies in the items of this examination list. For example, easing factors for pain were not included in the final examination. In addition, centralization of leg pain was included as an important feature, however the repeated active movements, needed to induce this change in pain response, were not

included. The study also took a uni-professional approach, and thus its findings may not be generalizable beyond physiotherapy; however, every effort was taken to ensure the widest range of opinion amongst the participating physiotherapists was sampled. It is envisaged that following the development of a classification system, based on this work, subsequent research will be multi-disciplinary. Finally, a further limitation is the concern, expressed by some authors, that the Delphi technique's consensus approach may prevent elaboration of the issues (14). However, attempts were made to reduce this limitation by allowing space for free text responses throughout the rounds of the Delphi.

This is the first study to establish the items of the clinical examination that a diverse sample of the UK physiotherapy profession agree should be included in an examination of NSLBP. The validity of these items will subsequently be evaluated in a multi-centre study being conducted across the UK. In this follow-up work the validity of each item on the list will be assessed for its measurement validity and contribute to the steadily growing body of evidence within this field. For over 10 years the sub-classification of NSLBP has been recommended as a priority for future research. It is likely that the establishment of a rigorously developed system that is valid, feasible and generalizable will allow a critical evaluation of the relative effectiveness of conservative treatments with the subgroups of patients most likely to benefit from them. Whilst the development of such a system will be challenging, the need for this advance in diagnostic practice is undoubted.

ACKNOWLEDGEMENTS

This work was funded by an Arthritis Research Campaign ICAC grant and a grant from the UK Department of Health, Research Capacity Development Agency.

REFERENCES

1. Foster NE, Thompson KA, Baxter GD, Allen JM. Management of nonspecific low back pain by physiotherapists in Britain and Ireland. Spine 1999; 24: 1332–1342.
2. van Tulder MV, Koes BW, Assendelft WJ, Bouter LM. The effectiveness of conservative treatment of acute and chronic low back pain. Den Haag: EMGO Institute; 1999.
3. Moffroid MT, Haugh LD, Henry SM, Short B. Distinguishable groups of musculoskeletal low back pain patients and asymptomatic control subjects based on physical measures of the NIOSH Low Back Atlas. Spine 1994; 19: 1350–1358.
4. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. Phys Ther 1995; 75: 470–485.
5. Doody C, McAteer M. Clinical reasoning of expert and novice physiotherapists in an outpatient setting. Physiotherapy 2002; 88: 258–268.
6. Fritz JM, George S. The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. Spine 2000; 25: 106–114.
7. Laslett M, Oberg B, Aprill CN, McDonald B. Centralization as a predictor of provocation discography results in chronic low back pain, and the influence of disability and distress on diagnostic power. Spine J 2005; 5: 370–380.
8. Petersen T, Laslett M, Manniche C, Ekdahl C, Jacobsen S. Diagnostic classification on non-specific low back pain. A new system integrating patho-anatomic and clinical categories. Physiother Theory Pract 2003; 19: 213–237.
9. 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.
10. Coste J, Paolaggi JB, Spira A. Classification of nonspecific low back pain. II. Clinical diversity of organic forms. Spine 1992; 17: 1038–1042.
11. Coste J, Paolaggi JB, Spira A. Classification of nonspecific low back pain. I. Psychological involvement in low back pain. A clinical, descriptive approach. Spine 1992; 17: 1028–1037.
12. McCarthy CJ, Arnall FA, Strimpakos N, Freemont AJ, Oldham JA. The biopsychosocial classification of non-specific low back pain: a systematic review. Phys Ther Rev 2004; 9: 17–30.
13. Kitzinger J. Introducing focus groups. Br Med J 1995; 311: 299–302.
14. Beech B. The delphi approach: recent applications in health care. Nurse Res 2001; 8: 38–47.
15. Krippendorff K. Content analysis: an introduction to its methodology. London: Sage; 1980.
16. Hicks CM. The Delphi technique. Research methods for clinical therapists: applied project design and analysis. Edinburgh: Churchill Livingstone; 1999, p. 243–249.
17. Moore AP, Petty NJ. Neuromusculoskeletal examination and assessment: a handbook for therapists. London: Churchill Livingstone; 2001.
18. McKenzie R. Low back pain. NZ Med J 1987; 100: 428–429.
19. McCarthy CJ. Spinal manipulative thrust technique using combined movement theory. Man Ther 2001; 6: 197–204.
20. Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Aust J Physiother 2003; 49: 89–97.
21. Laslett M, Oberg B, Aprill CN, McDonald B. Centralization as a predictor of provocation discography results in chronic low back pain, and the influence of disability and distress on diagnostic power. Spine J 2005; 5: 370–380.
22. Binkley J, Finch E, Hall J, Black T, Gowland C. Diagnostic classification of patients with low back pain: report on a survey of physical therapy experts. Phys Ther 1993; 73: 138–150.
23. Spitzer WO, LeBlanc FE, Dupuis M. Scientific approach to the assessment and management of activity related spinal disorders: report of the Quebec Task Force on Spinal Disorders. Spine 1987; 12: S16–S21.
24. Richardson C, Jull G, Hodges P, Hides J. Local muscle dysfunction in low back pain. Therapeutic exercise for spinal segmental stabilisation in low back pain. Edinburgh: Churchill Livingstone; 1999, p. 61–76.
25. Elvey RL. Physical evaluation of the peripheral nervous system in disorders of pain and dysfunction. J Hand Ther 1997; 10: 122–129.
26. McKenzie RA. The lumbar spine: mechanical diagnosis and therapy. Waikanae: Spinal Publications Ltd; 1981.
27. Edwards BC. Manual of combined movements, 1st edn. Edinburgh: Churchill Livingstone; 1992.
28. Maitland GD, Hengeveld E, Banks K, English K. Maitland's vertebral manipulation, Sixth edn. Oxford: Butterworth Heinemann; 2001.
29. Alvarez DJ, Rockwell PG. Trigger points: diagnosis and management. Am Fam Physician 2002; 65: 653–660.
30. Waddell G, Somerville D, Henderson I, Newton M. Objective clinical evaluation of physical impairment in chronic low back pain. Spine 1992; 17: 617–628.