Integration of subclassification strategies in randomised controlled clinical trials evaluating manual therapy treatment and exercise therapy for non-specific chronic low back pain: a systematic review

K V Fersum,[1] W Dankaerts,[2,3] P B O'Sullivan,[4] J Maes,[3] J S Skouen,[1,5] J M Bjordal,[1,6] A Kvåle[1]

[1]Section for Physiotherapy Science, Department of Public Health and Primary Health Care, University of Bergen, Norway
[2]Musculoskeletal Unit, Department of Rehabilitation Sciences, Faculty of Kinesiology and Rehabilitation Sciences, University College Limburg, KU Leuven, Leuven, Belgium
[3]College Limburg, Department of Health Care, AUHL-PHL, REVAL–Rehabilitation and Health Care Research Center, Hasselt, Belgium
[4]School of Physiotherapy, Curtin University, Bentley 6102, Western Australia, Australia
[5]The Outpatient Spine Clinic, Department of Physical Medicine and Rehabilitation, Haukeland University Hospital, Bergen, Norway
[6]Faculty of Health and Social Sciences, Bergen University College, Norway

Correspondence to
Kjartan Vibe Fersum, Section for Physiotherapy Science, Department of Public Health and Primary Health Care, University of Bergen, Kalfarveien 31, 5018 Bergen, Norway;
kjartan.fersum@isf.uib.no

Accepted 4 November 2009

ABSTRACT

Background There is lack of evidence for specific treatment interventions for patients with non-specific chronic low back pain (NSCLBP) despite the substantial amount of randomised controlled clinical trials evaluating treatment outcome for this disorder.

Hypothesis It has been hypothesised that this vacuum of evidence is caused by the lack of subclassification of the heterogeneous population of patients with chronic low back pain for outcome research.

Study design A systematic review.

Methods A systematic review with a meta-analysis was undertaken to determine the integration of subclassification strategies with matched interventions in randomised controlled clinical trials evaluating manual therapy treatment and exercise therapy for NSCLBP. A structured search for relevant studies in Embase, Cinahl, Medline, PEDro and the Cochrane Trials Register database, followed by hand searching all relevant studies in English up to December 2008.

Results Only 5 of 68 studies (7.4%) subclassified patients beyond applying general inclusion and exclusion criteria. In the few studies where classification and matched interventions have been used, our meta-analysis showed a statistical difference in favour of the classification-based intervention for reductions in pain ($p=0.004$) and disability ($p=0.0005$), both for short-term and long-term reduction in pain ($p=0.001$). Effect sizes ranged from moderate (0.43) for short term to minimal (0.14) for long term.

Conclusion A better integration of subclassification strategies in NSCLBP outcome research is needed. We propose the development of explicit recommendations for the use of subclassification strategies and evaluation of targeted interventions in future research evaluating NSCLBP.

Incidence of low back pain (LBP) in the Nordic population during a lifetime ranges from 60% to 65% and from 40% to 55% within a 12-month period.[1] Most cases (85%) are classified as non-specific because a definitive diagnosis cannot be achieved by current radiological methods.[2] This leaves a diagnostic and management vacuum.[3]

Classification systems (CS) can be defined as devices for sorting the complex elements of our reality into reasonable and logical entities.[4] More specifically, it has been proposed that a CS for non-specific chronic low back pain (NSLBP) should identify the underlying mechanisms driving the disorder within a biopsychosocial framework, enabling specific therapies to be applied so as to favourably influence the outcome of the disorder.[5] A recent study[6 7] demonstrated that there is a prevalent belief among primary care clinicians that NSLBP is a heterogenic condition, supporting that patients should be treated differently based on this heterogeneity. This practice is evident among clinicians, although there is little current evidence to support its validity.[7] It is considered that the heterogeneous group of NSLBP consists of several smaller homogenous subsets, with each subset being more likely to respond to a type of treatment unique to that classification.[8] 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 recommendation to establish methods of classification that will distinguish one subset from another.[9 10]

The Quebec Task Force Classification (QTFC) is considered by many the first "multidimensional" CS.[11] The QTFC consists of 11 subgroupings or categories and considers pathoanatomical diagnosis (specific or non-specific), red flags, signs and symptoms (area of pain referral), social factors, and the stage of disorder (acute, subacute or chronic). The system also acknowledges work status ("at work"/"not at work"). The QTFC was designed to assist in making clinical decisions, establishing a prognosis, evaluating the quality of care for patients with LBP and forming the basis for conducting scientific research.[11] However, it does not consider the underlying pain mechanism(s),[12] except for differentiating somatic from radicular pain. Hence, there is no subgrouping of NSLBP except based on pain area and no specific treatment is advocated for this large group of patients other than general exercise, therefore limiting its use for physiotherapy assessment and treatment.[13]

Randomised controlled clinical trials (RCTs) are, together with systematic reviews and meta-analysis, regarded to be the cornerstone of evidence-based medicine.[14 15] In a series of review articles in which RCTs were summarised, evaluating the efficacy of conservative treatments of LBP,[16–18] methodological flaws were commonly found by the reviewers. When undertaking a RCT, attempts should be made to include

a prognostic homogenous study group, which is likely to respond to the experimental intervention(s). However, in the available RCTs, it appears that heterogeneous study groups are often included.[19] This may hamper finding a treatment effect if, for instance, an intervention is effective only for a subset of the population. In this case the positive effect in this subgroup will be diluted because of the absence of effect in the complementary subgroups.

Exercise is widely used in the rehabilitation of patients with NSCLBP. However, no consensus exists as to the most effective programme design based on RCTs and systematic review.[20] The available evidence provides little guidance to clinicians who need to decide which interventions to implement for patients with NSCLBP. Furthermore, there is little basis on which to prefer manipulative therapy or exercise therapy.[17 21] A review by Liddle *et al*[20] highlighted the diversity of exercise programmes offered to patients with chronic low back pain (CLBP). Furthermore, no form of exercise has been shown to be more efficacious than another.[18]

Therefore, optimal treatment for patients with NSCLBP remains largely enigmatic.[22] It has been stated that caring for CLBP is one of the most difficult and unrewarding problems in clinical medicine[23] because no treatment has been shown to be clearly effective.[17 24 25] It has been hypothesised that the lack of evidence for managing NSCLBP is a result of the vast majority of RCTs broadly defining heterogeneous populations.[17 24] Furthermore, it has been stated that without subclassification, research into NSCLBP will be unlikely to provide useful insight.[26]

Therefore, the aim of this study was to review the literature on RCTs evaluating manual therapy treatment and exercise therapy for patients with NSCLBP. More specific, the aims were to investigate the level of integration of subclassification in these RCTs and to summarise the effects of the studies that had subclassified and matched treatments accordingly based on an meta analysis (MA) To our best knowledge, no study has systematically reviewed this.

METHODS

Inclusion and exclusion criteria

The process of a systematic review involves thorough investigation aimed at identifying all studies on a specific topic.[27] The flowchart in fig 1 is a summary of the multilevel reviewing process.

In this review, only RCTs that evaluated the effect of manual therapy treatment or exercise therapy interventions for adult patients with NSCLBP, where symptom duration exceeded 12 weeks, were included (level 1). Medical therapy, surgery, acupuncture and electrotherapy are interventions that were excluded. The second part of the multilevel evaluation process (level 2) was to review the included papers with regards to the patient sample of each RCT. This part was deemed essential to ensure that the studies included actually dealt with NSCLBP disorders. This process involved looking specifically at what exclusion criteria the study had chosen with regards to red flag disorders (cancer, inflammatory disorders, infection or fractures), specific pathology (spondylolisthesis, disc herniation with radicular pain, degenerative disc with Modic changes, central or foraminal stenosis), pregnancy and surgery. Whether participants with a known psychiatric disorder were excluded was also noted at this level. Level 3 reviewed if the RCT's inclusions criteria were made in relation to the time frame (acute, subacute, chronic), age of participants, area of pain, level of pain, level of disability, psychosocial status, work status and finally whether there was any compensation involved or litigation pending. These two levels (levels 2 and 3) were important to qualitatively the level of heterogeneity of the NSCLBP subjects included in the RCTs. Level 4 investigated specifically whether classification strategies had been attempted and whether treatment was matched to subgroups based on classification. Any attempt to subgroup patients based on specific characteristics, physical or psychological, was accepted. The subclassification could be prospective or retrospective. The final part level 4 was to do an MA of these studies where subclassification had been attempted or used, to determine the effect of matched interventions. The main outcomes for the MA were pain intensity (visual analogue scale (VAS)) and disability (Roland Morris or Oswestry). A further description of these studies and their validation can be found in table 3.

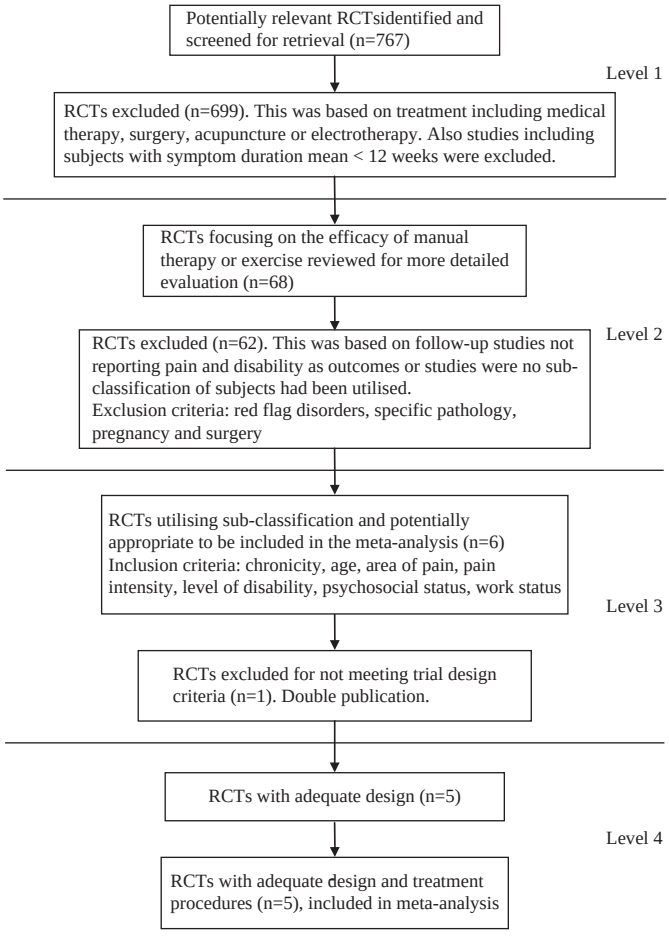

Figure 1 Summary of the multilevel reviewing process.

Search strategy

The electronic databases Medline (1966 to December 2008), Cinahl (1982 to December 2008), Embase (1988 to December 2008) and the Cochrane Central Register of Controlled Trials (fourth Quarter 2008) were searched via Ovid to identify all relevant trials. This is in line with the recommendations from studies of Minozzi *et al*[28] and Woods and Trewheellar (1998)[29] where both Medline and Embase are suggested to be used to ensure a comprehensive literature search because the overlap between these two databases is small. In addition, we followed minimum search strategy as suggested by Van Tulder *et al*.[30] The search was limited to articles in English and pertaining to human subjects. All reference lists of trials were

identified through electronic searching with both MeSH terms and single terms and searched recursively until no more trials were identified. Keywords and combinations were low back pain, chronic pain AND manipulative medicine, kinesiotherapy, exercise therapy AND randomised controlled trial, RCT, clinical trial. The next phase of the search strategy involved manual selection of the obtained search results.

Assessment of study quality

The methodological quality of the five studies included in the MA was assessed using the PEDro scale. This scale is an 11-item scale designed to assess the methodological quality of RCTs (table 1). Points are only awarded when a criterion is clearly satisfied and reported. If on a literal reading of the trial report a criterion was not satisfied or reported, then no points are awarded (see http://www.pedro.fhs.usyd.edu.au/scale_item.html for further details). For this review it was decided that using the scoring taken from PEDro database would be the most objective way of scoring these papers because this scoring is not influenced by or related to any particular research purpose (table 1).

Data extraction

Two independent reviewers (JM and KVF) conducted searches and assessed the trials for eligibility. The titles and abstracts were then further screened for suitability for inclusion (level 1). Disagreements were resolved with a consensus meeting between reviewers. Many of the studies had included a mix of patients with acute, subacute LBP and CLBP, as well as specific LBP disorders. After the screening, data about age, chronicity, area of pain, level of pain/disability, psychosocial status, work status and any form of compensation of the participants from the included papers were extracted (level 2). Studies that attempted a specific subclassification strategy beyond general inclusion and exclusion criteria in line with the criteria described in the methods were included (level 3). From these studies, short-term and long-term data for the outcomes of interest (pain and disability), means and SD were extracted. With these data, an MA was performed at the final level (level 4).

Data analysis

For levels 2 and 3, results were plotted and frequencies were calculated on a percentage basis. Trials were assessed for clinical heterogeneity with respect to their inclusion and exclusion criteria. To do an MA of the effectiveness, we extracted the group means and SD for each comparison using the outcome measure (pain and disability) in these studies that had attempted subclassification (level 4). In two of the studies included in the MA[31 32] where subclassification had been made based on the Multidimensional Pain Inventory, the data were extracted and plotted for each of the different subgroups to show the effect for each of these independently (see figs 3–5). In cases of missing data where studies failed to report SD, we calculated SD from other variance data or imputed a reasonable SD value.[33] Pain intensity on a 100-mm VAS was defined as the pooled estimate of the difference in change between the means of the treatment and the placebo/control groups, weighted by the inverse of the pooled SD of change for each study, that is, weighted mean difference of change between groups. The variance was calculated from the trial data and with 95% confidence interval (CI) in millimetres on VAS.

Because of the possibility of the outcome measurement by different disability scales, these were defined as unit less pooled estimate of the difference in change between the mean of treatment and control group, weighted by the inverse of the pooled SD of change for each study, that is, standardised mean difference of change between groups using Review Manager (RevMan) V.5.0.18 2008 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). The variance was calculated from the trial data with 95% CI. Results were considered significant if $p<0.05$. For reasons relating to generalisability and given this review investigated NSCLBP, we considered it appropriate to conduct MA both for short-term and long-term outcomes. Four MA were performed (short- and

Table 1 Trial quality assessed by the PEDro scale with the criteria 1–11*

Trial	1	2	3	4	5	6	7	8	9	10	11	Total
Gudavalli *et al* (2006)	+	+	+	+	−	−	−	−	−	+	+	5/10
Riipinen *et al* (2005)	+	+	+	−	−	−	−	−	−	+	+	4/10
Vollenbroek-Hutten *et al* (2004)	+	+	−	+	−	−	+	+	+	+	+	7/10
Petersen *et al* (2002)	+	+	+	+	−	−	−	+	+	+	+	7/10
Snook *et al* (1998)	+	+	−	+	−	−	−	−	+	+	−	4/10

Trials listed in descending order according to PEDro quality score.
*(1) Eligibility criteria were specified. (This criterion influences external validity, but not the internal or statistical validity of the trial. It has been included in the PEDro scale so that all items of the Delphi scale are represented on the PEDro scale. This item is not used to calculate the PEDro score.)
(2) Subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received).
(3) Allocation was concealed.
(4) The groups were similar at baseline regarding the most important prognostic indicators.
(5) There was blinding of all the subjects.
(6) There was blinding of all therapists who administered the treatment.
(7) There was blinding of all assessors who measured at least one key outcome.
(8) Measurements of at least one key outcome were obtained from >85% of the subjects initially allocated to groups.
(9) All subjects for whom outcome measurements available received the treatment or control condition as allocated, or where this was not the case, data for at least one key outcome were analysed by "intention to treat".
(10) The results of between-group statistical comparisons are reported for at least one key outcome.
(11) The study provides both point measurements and measurements of variability for at least one key outcome.

Table 2 Studies integrating subclassification strategies and their validity

Author (N)	Description	Does it reflect the biopsychosocial model?	Has it been validated for NSCLBP?	Is it known to be reliable?	Is the intervention matched?
Gudavalli *et al* (2006) (N=235)	Subgroup categorisation was based on physical examination, history and clinician's assessment of severity and recurrence of pain episode (1) Subjects were first divided according to the presence of radiculopathy (yes/no). Evidence of radiculopathy was included because the prognosis for patients with radiculopathy may be different from those without [AU: Please check edited table entry if correct.] (2) Subjects were then divided according to the level of severity (3) Then, division of subject were made according to whether the term *chronic* referred to continually present pain or recurrent pain	No	No	No	Yes
Riipinen *et al* 2005 (N=204)	The Multidimensional Pain Inventory (MPI) is a method for identifying patient subgroups to aid in predicting rehabilitation outcome. Three patient profiles can be derived empirically from it: (1) Adaptive copers (AC) report lower levels of pain severity, interference with activities and affective distress and report greater perceptions of life control and activity level (2) Those interpersonally distressed (ID) are characterised by lower levels of social support and lower scores on receiving solicitous and distracting responses from their significant others (3) Dysfunctionals (DYS) are distinguished by a higher level of pain severity, marked interference of pain in everyday life, high affective distress, low perception of life control and low levels of activity. The uniqueness of these profiles has been validated with various patient sample measurement instruments and with different translations of the MPI. It is suggested that patients with distinct profiles may respond differently to standard treatment. They may also gain an advantage from different types of intervention targeted to their specific needs, for instance, DYS patients may benefit from interventions focusing on psychosocial distress and stress management in addition to physical assessment, whereas ID patients may benefit from a specific focus on interpersonal skills and problem-solving. AC patients, on the other hand, may benefit from a focus only on somatic disorders with no psychosocial components. Patients were randomly assigned either to a combined manipulation, exercise and physician consultation group (called the combination group) or to physician consultation-alone group (called the consultation group); the latter served as the control. They used a Finnish translation of the MPI, SIMPI, which was shown in an earlier study as effective in producing the three cluster profiles introduced by Turk and Rudy[37]	Yes	No	Yes	No
Vollenbroek-Hutten *et al* 2004 (N=142)	Based on the MPI-DLV24 and lumbar dynamometry results, both measured at inclusion, patients were divided into the following subgroups: (1) Patients with performances lower than healthy subjects but with consistent test behaviour (expected performance) (2) Patients with performances comparable to those of healthy subjects and consistent test behaviour (normal performance) (3) Patients with inconsistent test behaviour, meaning that their performance is not maximal and that the assessment is (probably) not valid. Dependent on the number of inconsistencies, this is called grey zone or submaximal performance	Yes	No	Yes	No

Continued

Table 2 Continued

Author (N)	Description	Does it reflect the biopsychosocial model?	Has it been validated for NSCLBP?	Is it known to be reliable?	Is the intervention matched?
Petersen *et al* 2002 (N= 260)	The McKenzie treatment was planned individually after an initial assessment according to the principles described by Robin McKenzie. This follows an algorithm that leads to the simple classification of spinal-related disorders. It is based on a consistent "cause and effect" relationship between historical pain behaviour as well as the pain response to repeated test movements, positions and activities during the assessment process. A systematic progression of applied mechanical forces uses pain response to monitor changes in motion/function. The McKenzie classification of spinal pain provides reproducible means of separating patients with apparently similar presentations into definable subgroups (syndromes) to determine appropriate treatment. McKenzie has named these three mechanical syndromes: (1) Postural: end-range stress of normal structures (2) Dysfunction: end-range stress of shortened structures (scarring, fibrosis, nerve root adherence) (3) Derangement: anatomical disruption or displacement within the motion segment. Each distinct syndrome is addressed according to its nature with mechanical procedures using movement and positions. The Derangement syndrome where the phenomenon of "centralisation" occurs is most common. The treatment emphasises education and active patient involvement in the management of their treatment to decrease pain and restore function and independence	No	No	Yes	Yes
Snook *et al* 1998 (N=85)	Data collected during the initial visit to the Research Centre were used to group the subjects into four dichotomous subgroups: (1) Age (30–45 and 46–60 years) (2) Sex (3) Location of pain (back pain only and back and leg pain below the knee) (4) Psychological overlay (high and low). Psychological overlay was determined from the Modified Zung Depression Index and the Modified Somatic Perception Questionnaire, using the criteria proposed by Main. Subjects were divided into the 16 combinations of the four dichotomous subgroups and then randomised within each of the 16 combinations. Employment and occupational data were collected during the third visit. These data were used to allocate subjects into physically heavy, moderate and light work groups	Yes	No	No	Yes

long-term for both pain and disability). Because the treatment periods and long-term follow-up varied, end of treatment and long-term follow-up of 36–52 weeks were chosen as measurement points. The statistical heterogeneity (genuine differences underlying the results of the trials in the review) of the results of the trials was measured using the quantity I^2. Because using the p value as a measure for heterogeneity ($p<0.10$) has been known to be poor at detecting true heterogeneity among studies as significant, it has been suggested that the quantity I^2 should be used instead.[34] This value can be calculated as $I^2=100\%(Q-df)/Q$, where Q is Cochran's heterogeneity statistic and df is the degree of freedom (where n is the number of trials and therefore degrees of freedom equals number of studies minus 1). The Cochran's Q is computed by summing the squared deviations of each trial's estimate and a p value from the overall meta-analytical estimate and a p value obtained by comparing the statistic with a X^2 distribution with $k-1$ degrees of freedom (where k is the number of studies).[34] Trials in the MA were considered to have low statistical heterogeneity if I^2 <25 %, and in such instances, a fixed-effect model should be used. This assumes that the true effect of treatment is the same value in every trial. In contrast, random-effects MA model assumes that the effects being estimated in the different studies are not identical but follow a similar distribution.

RESULTS

Search results

A total of 767 RCTs published between 1982 and December 2008 that administered conservative treatment for LBP were identified and screened (level 1). However, only 68 studies

had focused on the efficacy of manual therapy or exercise and reported outcomes based on levels of disability and pain in subjects with NSLBP (level 2). At this level we registered that 11 studies (16.2%) did not report the exclusion of red flags or specific spinal pathology as defined. Fourteen RCTs (20.6%) did not exclude or subgroup subjects with nerve root irritation/pathology (as defined at level 2). Twenty-nine studies (42.6%) had not listed pregnancy as an exclusion criterion. Twenty-three studies (33.8%) did not exclude subjects if they had undergone surgery for their LBP. We also found that seven studies (10.3%) had not specified a time frame for LBP duration in their inclusion criteria. Fourteen studies (20.6%) had included patients with <3 months durations of symptoms. However, in 21 of these studies, >90% of the patients had pain lasting >3 months. The remaining 47 studies (69.1%) had specified in the inclusion that symptom durations had to be >3 months. For age, most of the studies (56 (82.3%)) specified that the patients had to be between 18 and 65 years old. Only one study included patients >65 years old and 11 studies (16.2%) did not specify age. Psychosocial status was only specified in 1 study (1.5%), work status in 14 (20.6%) and compensation in 11 (16.2%). Eleven studies (16.2%) specified pain intensity level and 23 (33.8%) specified disability level as an outcome.

At level 3, only 6 (8.8%) of the 68 studies had performed some form of subclassification according to definitions described previously in the methods. One of these studies was a double publication,[35] and therefore only the original study[36] was included in the next level. These studies were therefore included in the MA (level 4), providing information on altogether 432 participants for disability and 359 participants for pain. The data from the MA show a statistical difference in favour of the classification-based intervention for reductions in pain (p=0.004) and disability (p=0.0005), both for short-term and long-term reduction in pain (p=0.001). Disability did not reach statistical significance (p=0.07) for long-term outcome. Effect sizes ranged from moderate (0.43) at short term to minimal (0.14) for long term. Table 2 shows a description of the included papers and their subclassification strategy.

DISCUSSION

This systematic review found a very low level of integration of subclassification strategies in RCTs for NSCLBP. On levels 1 and 2, a number of studies fail to apply the most essential and basic form of classification. Despite this, it is part of a well-accepted diagnostic "triage" process of classifying LBP.[38 39] Eleven studies (16.2%) did not mention an attempt to exclude patients with red flag or any specific spinal pathology, and 14 studies (20.6%) did not exclude nor identify patients with nerve root irritation/pathology, although the studies aimed to target NSCLBP.

Twenty-nine of the 68 RCTs (42.6%) did not exclude pregnant subjects, although it is well accepted that pregnancy-related pelvic girdle pain represents a small but specific subgroup of musculoskeletal disorders.[40] The mechanisms by which pregnancy can affect the lumbar spine and pelvic girdle complex is not fully understood.[41] Multiple mechanisms have been suggested along with specific tailored treatment.[40 42 43]

We found it surprising that 23 studies (33.8%) did not exclude subjects if they had undergone surgery for their LBP. Patients eligible for surgery presumably had a specific pathology that was responsible for their pain and/or disability. On one hand, it could be argued that these patients should be excluded based on the initial diagnostic triage process of classifying LBP.[39] On the other hand, it could also be said that if they still had pain after surgery, it is likely that the lesion operated on was in fact not the cause of their initial pain and therefore possibly these patients should have been classified as NSCLBP presurgical. However, as complications after failed back surgery, scar tissue formation and tissue sensitisation occurs in 10–40% and the incidence of reoperation is one in four patients.[38 44] This group of patients is likely to be different from non-surgical patients and therefore potentially makes the population in these RCTs more heterogenic.

Traditionally, CLBP has been defined as LBP, where the duration of pain has lasted >3 months or occurs episodically within a 6-month period.[45]

Reviewing the literature for NSCLBP, we encountered broad definitions for the term *chronic*. Of the 68 studies, 21 (30.9%) violated against this basic selection criterion by including patients with <12 weeks of symptom duration. Of these 21 studies, 7 RCTs did not list any time frame and 14 RCTs included patients with <3 months of pain.

The biopsychosocial model of chronic pain emphasises the interplay of biological and psychosocial factors in the development, expression and maintenance of pain.[46] One study[47] reported the psychosocial status of their patients in the inclusion criteria, whereas several RCTs were evaluating multidimensional interventions (with combined psychological-orientated interventions and exercise). However, applying a form of psychotherapy on patients who are not screened for psychological issues appears to be a major limitation, given that many subjects with CLBP may not have significant psychological factors driving their disorder.

Only 7.4% (5/68) of the RCTs included in the systematic review used a subclassification strategy beyond general inclusion and exclusion criteria (level 3). The MA (level 4) indicated there was a statistically significant effect in favour of these interventions where subclassification strategies had been used. This was for short-term (end of treatment) and long-term (36–52 weeks follow-up) pain intensity levels (figs 2 and 3). For disability reduction there was also a statistically significant effect in favour of short term and no statistical significance (p=0.07) was reached for long term (36–52 weeks) (figs 4 and 5). These data should however be interpreted with caution as the numbers are insufficient to definitively quantify the effect of subclassification strategies for the treatment of NSCLBP. There is a need for more high-quality RCTs with similar comparisons.[48] The quality of these five studies also needs to be considered as two out of five scored <5 on the PEDro scale (table 1). There are a number of additional factors that need to be considered following the results of this MA.[49] These will be discussed later.

Most CS to date that relate to LBP have focused on a single dimension rather than on considering all dimensions of LBP within a biopsychosocial perspective.[50] It can therefore be argued that for a CS to be clinically useful, it should be based on identifying the underlying mechanism(s) driving the pain disorder, thereby guiding targeted interventions.[5 10] This also requires a CS to follow a thorough validation process involving multiple steps with different stages dealing with different criteria as suggested by Dankaerts *et al*.[51]

Three of the five studies included[31 32 52] used a CS method that incorporates the biopsychosocial model. Riipinen *et al*[31] and Vollenbroek-Hutten[32] used the Multidimensional Pain Inventory (MPI) (see table 3) to subclassify. Previous studies have shown that using this multiaxial assessment gives more consistent results[53] than studies using single aspects (physical, psychological or social) in combination.[54 55] However, the validity of these studies including a biopsychosocial model can

be questioned as neither matched treatment according to the MPI profile nor the subclassification analysis was performed retrospectively. In addition, the study by Snook *et al*[52] incorporated both the biopsychosocial and psychosocial dimensions when subgrouping their patients. Although the analysis in these three studies may give important information about responders to certain treatments, neither uses subclassification strategies to tailor the treatment based on the proposed underlying mechanism(s) of the LBP disorder. Therefore, it is not possible to validate a CS using this approach for the management of NSCLBP. The two other studies[36 56] reviewed in level 4 of the reviewing process are not fully validated for NSCLBP either. The McKenzie CS used by Petersen *et al*[57] have been tested for acute and subacute LBP patients in several studies with conflicting results.[58] Although it seems to have some validity for these patient groups, its efficacy and validation for NSCLBP are still not established.[59] In the study by Gudavalli *et al*,[36] a form of subclassification was made retrospectively to assess responders to flexion-distraction treatment or active exercise. Their evaluation of subgroups (table 2), although it has some validation from the QTFC, cannot claim to be fully validated for NSCLBP.

In three of the five studies included in the MA,[31 32 56] reliability of the subgroups in the CS have been tested. The McKenzie-based CS was tested by Petersen *et al*,[60] and the overall percentage of agreement between examiners was 72% and the κ coefficient was 0.62 (95% CI 0.50 to 0.74). Previous research on the MPI has demonstrated that it had good reliability[61] and is sensitive to change.[62] In the two other studies,[52] no reliability data are reported based on the different subgroups.

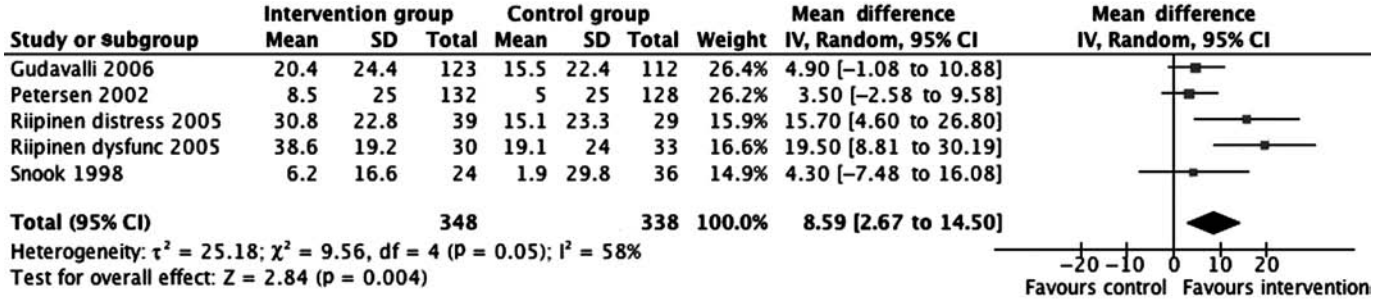

Figure 2 End of treatment results for subclassified intervention measured as the weighted mean difference pain reduction on 100-mm visual analogue scale. Trials are subgrouped based on the different subclassifications and combined results are shown as total on the bottom of the table. Plots on the right side of the middle line indicate that the effect of the intervention (subclassification) is superior to the control intervention.

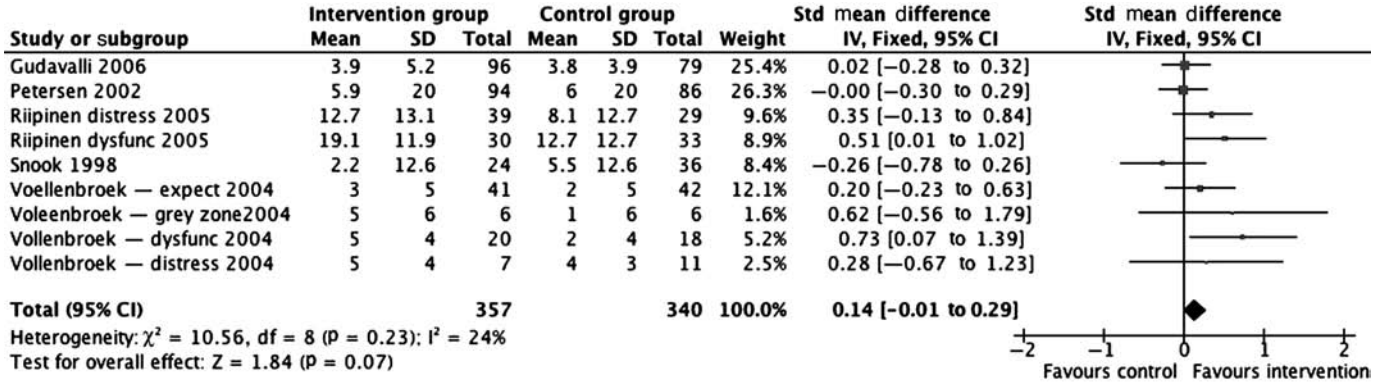

Figure 3 Follow-up 36–52 weeks results for subclassified intervention measured as the weighted mean difference pain reduction on 100-mm visual analogue scale. Trials are subgrouped based on the different subclassifications and combined results are shown as total on the bottom of the table. Plots on the right side of the middle line indicate that the effect of the intervention (subclassification) is superior to the control intervention.

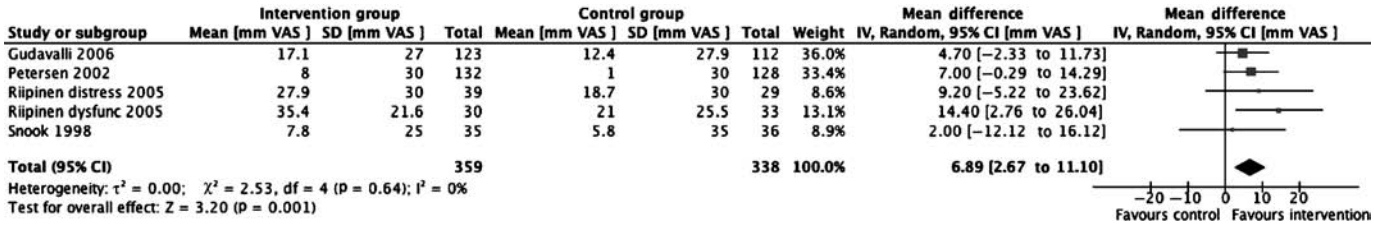

Figure 4 End of treatment results for subclassified intervention measured as the standardised mean difference of change in disability between groups. Trials are subgrouped based on the different subclassifications and combined results are shown as total on the bottom of the table. Plots on the right side of the middle line indicate that the effect of the intervention (subclassification) is superior to the control intervention.

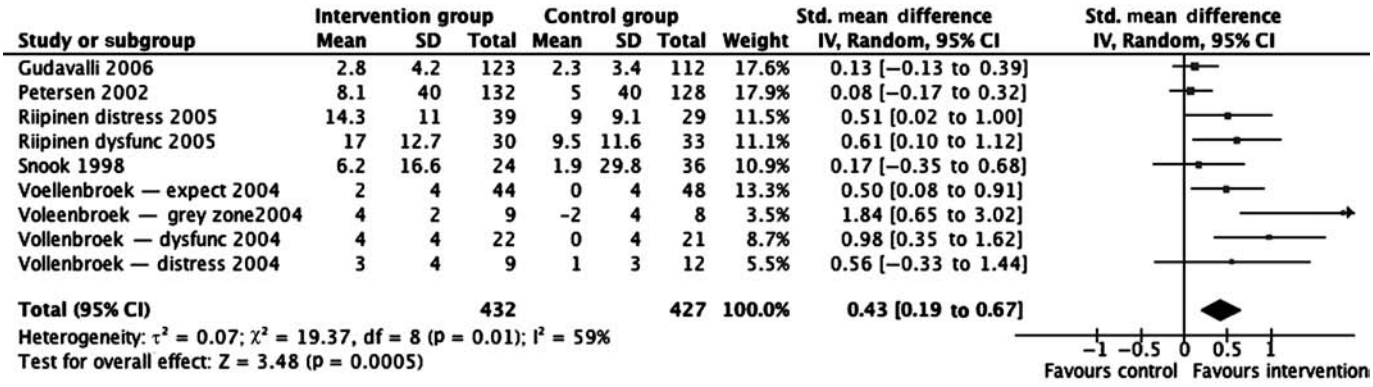

Figure 5 Follow-up 36–52 weeks treatment results for subclassified intervention measured as the standardised mean difference of change in disability between groups. Trials are subgrouped based on the different subclassifications and combined results are shown as total on the bottom of the table. Plots on the right side of the middle line indicate that the effect of the intervention (subclassification) is superior to the control intervention.

Two of the five studies[52 56] matched their specific intervention to subgroups based on proposed theoretical frameworks. Petersen *et al*[56] used a McKenzie-based[63] CS. This system is based on information from history taking and symptom response to generated loading of the lumbar spine (table 3). However, the system only has a pathoanatomical basis and lacks clear guidelines for management. In the study by Snook *et al*,[52] the intervention consisted of instruction in the control of early morning lumbar flexion compared with sham treatment (six exercises). Data were analysed based on four different subgroups (table 3). The subjects were given a back scratcher and a reacher. After 6 hours of not bending, usual activities were allowed, but extreme bending should be avoided. The theory of Snook *et al*[52] was based on studies by Adams *et al*[64] that showed an increased risk of injury when bending was performed early in the morning. Their protocol therefore may be an initial attempt to target an assumed underlying mechanism for LBP, namely, the disc being more vulnerable and prone to injury in the morning because of increased water content. However, it is well established that CLBP is a multidimensional problem likely consisting of a combination of pathoanatomical, neurophysiological, physical and psychosocial factors.[59 65 66] Thus, it is unlikely that a general intervention such as "preventing early morning flexion" could target all these underlying mechanisms and resolve the complexity of NSCLBP.

Because of the shortcomings of the current classification models used in the RCTs under review, it is clear that NSCLBP disorders require further integration of classification strategies based on a biopsychosocial construct. There is growing evidence to support that subgroups do exist within both the physical and psychosocial domains.[10 67] There are a number of key clinical indicators regarding pain area and behaviour of symptoms, which provide important insight into the different mechanisms underlying and driving LBP disorders, in combination with a comprehensive physical examination allowing classification to be made.[68]

CONCLUSION

RCTs evaluating manual therapy treatment and exercise therapy in patients with CLBP using a CS approach and matched treatments are very limited to non-existing. Although there is still a need for studies developing more reproducible and accurate characterisation of subgroups of patients with CLBP, we recommend that for future RCTs investigating specific interventions for CLBP, available valid and reliable subclassification strategies operating within a biopsychosocial framework should be integrated. Well-defined inclusion and exclusion criteria should be used, interventions should target the underlying mechanism(s) and minimum standards for evaluation of outcome and long-term follow up should be established. Otherwise, research into NSCLBP will be unlikely to provide useful insight[26] into more effective management for this complex disorder.

Funding The Norwegian Fund for Postgraduate Training in Physiotherapy (Oslo, Norway).

Competing interests None.

Ethics approval This study was conducted with the approval of the regional committees for medical and health research ethics (REK).

Provenance and peer review Not commissioned; externally peer reviewed.

Patient consent Not obtained.

REFERENCES

1. **Leboeuf-Yde C**, Klougart N, Lauritzen T. How common is low back pain in the Nordic population? Data from a recent study on a middle-aged general Danish population and four surveys previously conducted in the Nordic countries. *Spine* 1996;**21**:1518–25; discussion 1525–6.
2. **Dillingham T**. Evaluation and management of low back pain: and overview. *State Art Rev* 1995;**9**:559–74.
3. **Leboeuf-Yde C**, Lauritsen JM, Lauritzen T. Why has the search for causes of low back pain largely been nonconclusive? *Spine* 1997;**22**:877–81.
4. **Malterud K**, Hollnagel H. The magic influence of classification systems in clinical practice. *Scand J Prim Health Care* 1997;**15**:5–6.
5. **O'Sullivan PB**. Diagnosis and classification of lumbo-pelvic pain disorders. *Man Ther* 2005;**10**:242–55.
6. **Kent P**, Keating J. Do primary-care clinicians think that non-specific low back pain is one condition? *Spine* 2004;**29**:1022–31.
7. **Kent P**, Keating JL. Classification in non-specific low back pain: what methods do primary care clinicians currently use? *Spine* 2005;**30**:1433–40.
8. **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**:106–14.
9. **Riddle DL**. Classification and low back pain: a review of the literature and critical analysis of selected systems. *Phys Ther* 1998;**78**:708–37.
10. **Turk DC**. The potential of treatment matching for subgroups of patients with chronic pain. *Clin J Pain* 2005;**21**:44–55.
11. **Spitzer W**. Scientific approach to the assessment and management of activity related spinal disorders. *Spine* 1987;**7**:S1–55.
12. **Dankaerts W**, O'Sullivan PB, Straker LM, *et al*. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. *Man Ther* 2006;**11**:28–39.
13. **Padfield B**. Use of an outcome measurement system to answer a clinical question: is the Quebec task force classification system useful in an outpatient setting? *Physiotherapy Canada* 2002:254–60.

14. **Craig JC**, Irwig LM, Stockler MR. Evidence-based medicine: useful tools for decision making. *Med J Aust* 2001;**174**:248–53.
15. **Devereaux PJ**, Yusuf S. The evolution of the randomised controlled trial and it role in evidence-based decision making. *J Intern Med* 2003;**254**:105–13.
16. **Assendelft WJ**, Morton SC, Yu EI, *et al.* Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Ann Intern Med* 2003;**138**:871–81.
17. **Hayden JA**, Van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain 1. *Ann Intern Med* 2005;**142**:776–85.
18. **van Tulder M**, Malmivaara A, Esmail R, *et al.* Exercise therapy for low back pain: a systematic review within the framework of the Cochrane collaboration back review group. *Spine* 2000;**25**:2784–96.
19. **Bouter LM**, van Tulder MW, Koes BW. Methodologic issues in low back pain research in primary care. *Spine* 1998;**23**:2014–20.
20. **Liddle SD**, Baxter GD, Gracey JH. Exercise and chronic low back pain: what works? *Pain* 2004;**107**:176–90.
21. **Ferreira ML**, Ferreira PH, Latimer J, *et al.* Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: a randomised trial. *Pain* 2007;**131**:31–7.
22. **Van Tulder MW**, Koes B, Bouter LM. Conservative treatment of acute and chronic non-specific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 1997;**22**:29.
23. **Leclere H**, Beaulieu G, Bordage A, *et al*. Why are clinical problems difficult? General practitioners' opinion concerning 24 clinical problems. *CMAJ* 1990;**143**:1305–15.
24. **Assendelft WJJ**, Morton SC, Yu EI, *et al*. Spinal manipulative therapy for low back pain. *Cochrane Database Syst Rev* 2004;**1**:CD000447.
25. **Mannion AF**, Müntener M, Taimela S, *et al.* Comparison of three active therapies for chronic low back pain: results of a randomised clinical trial with one-year follow-up. *Rheumatology (Oxford)* 2001;**40**:772–8.
26. **Leboef-Yde C**, Manniche C. Low back pain: time to get off the treadmill. *J Manipulative Physiol Ther* 2001;**24**:63–66.
27. **McNeely ML**. A systematic review of physiotherapy for spondylosis and spondylolisthesis. *Man Ther* 2003;**8**:80–90.
28. **Minozzi S**, Pistotti V, Forni M. Searching for rehabilitation articles on Medline and Embase. An example with cross-over design. *Arch Phys Med Rehabil* 2000;**81**:3.
29. **Woods D**, Trewheellar K. Medline and Embase complement each other in literature searches. *BMJ* 1998;**316**:1166.
30. **Van Tulder M**, Furlan A, Bombardier C, *et al.* Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine* 2003;**8**:10.
31. **Riipinen M**, Niemistö L, Lindgren KA, *et al*. Psychosocial differences as predictors for recovery from chronic low back pain following manipulation, stabilising exercise and physician consultation or physician consultation alone. *J Rehabil Med* 2005;**37**:152–8.
32. **Vollenbroek-Hutten MM**, Hermens HJ, Wever D, *et al*. Differences in outcome of a multidisciplinary treatment between subgroups of chronic low back pain patients defined using two multiaxial assessment instruments: the multidimensional pain inventory and lumbar dynamometry. *Clin Rehabil* 2004;**18**:556–69.
33. **Furukawa TA**, Barbui C, Cipriani A, *et al.* Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 2006;**59**:7–10.
34. **Higgins JP**, Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60.
35. **Cambron JA**, Gudavalli MR, Hedeker D, *et al*. One-year follow-up of a randomised controlled clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. *J Altern Complement Med* 2006;**12**:659–68.
36. **Gudavalli MR**, Cambron JA, McGregor M, *et al*. A randomised controlled clinical trial and subgroup analysis to compare flexion-distraction with active exercise for chronic low back pain. *Eur Spine J* 2006;**15**:1070–82.
37. **Turk DC**, Rudy TE. Toward an empirically derived taxonomy of chronic pain patients: integration of psychological assessment data. *J Consult Clin Psychol* 1988;**56**:233–8.
38. **Spitzer WO**, Leblanc F, Dupuis M. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. *Spine* 1987;**12**:S1–59.
39. **Materson RS**. The AHCPR practice guidelines for low back pain. *Bull Rheum Dis* 1996;**45**:6–8.
40. **O'Sullivan PB**, Beales DJ. Diagnosis and classification of pelvic girdle pain disorders—Part 1: a mechanism based approach within a biopsychosocial framework. *Man Ther* 2007;**12**:86–97.
41. **Kristiansson P**, Svardsudd K, Schoultz B. Back pain during pregnancy: a prospective study *Spine* 1996;**21**:702–9.
42. **Petersen LK**, Hvidman L, Uldbjerg N. Normal serum relaxin in women with disabling pelvic pain during pregnancy. *Gynecol Obstet Invest* 1994;**38**:3.
43. **Stuge B**, Laerum E, Kirkesola G, *et al.* The efficacy of a treatment program focusing on specific stabilising exercises for pelvic girdle pain after pregnancy: a randomised controlled trial. *Spine* 2004;**29**:351–9.
44. **Atlas SJ**, Keller RB, Wu YA, *et al.* Long-term outcomes of surgical and non-surgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. *Spine* 2005;**30**:927–35.
45. **Krismer M**, van Tulder M; Low Back Pain Group of the Bone and Joint Health Strategies for Europe Project. Strategies for prevention and management of musculoskeletal conditions. Low back pain (non-specific). *Best Pract Res Clin Rheumatol* 2007;**21**:77–91.
46. **Hanley MA**, Jensen MP, Ehde DM, *et al.* Psychosocial predictors of long-term adjustment to lower-limb amputation and phantom limb pain. *Disabil Rehabil* 2004;**26**:882–93.
47. **Kole-Snijders AM**, Vlaeyen JW, Goossens ME, *et al.* Chronic low-back pain: what does cognitive coping skills training add to operant behavioral treatment? Results of a randomised clinical trial. *J Consult Clin Psychol* 1999;**67**:931–44.
48. **Egger M**, Davey Smith G, Schneider M, *et al.* Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.
49. **Egger M**, Smith GD. Misleading meta-analysis. *BMJ* 1995;**310**:752–4.
50. **Ford J**, Story I, McMeeken J. A systematic review on methodology of classification system research for low back pain. *Musculoskeletal Physiotherapy Australia 13th Biennial Conference*, Sydney, Australia; 27-30 November 2003.
51. **Dankaerts W**, O'Sullivan PB, Burnett AF, *et al.* What precedes RCTs? A new model for clinical research into NS-CLBP. In 8th International Federation of Orthopaedic Manipulative Therapist' Conference, Cape Town, South Africa, 2004.
52. **Snook SH**, Webster BS, McGorry RW. The reduction of chronic, non-specific low back pain through the control of early morning lumbar flexion: 3-year follow-up. *J Occup Rehabil* 2002;**12**:13–9.
53. **Talo S**, Rytökoski U, Puukka P. Patient classification, a key to evaluate pain treatment: a psychological study in chronic low back pain patients.Volvo award in clinical science. *Spine* 1992;**17**:998–1011.
54. **Fredrickson B**, Trief P, VanBeveren P, *et al.* Rehabilitation of the patient with chronic back pain; a search for outcome predictors. *Spine* 1988;**13**:351–3.
55. **Haazen IWCJ**, Vlayen JWS, Kole-Snijders A. Behavioural rehabilitation of chronic low back pain: searching for predictors of treament outcome. *J Rehabil Sci* 1994;**7**:12.
56. **Petersen T**, Kryger P, Ekdahl C, *et al.* The effect of McKenzie therapy as compared with that of intensive strengthening training for the treatment of patients with subacute or chronic low back pain: a randomised controlled trial. *Spine* 2002;**27**:1702–9.
57. **Petersen T**, Thorsen H, Manniche C, *et al.* Classification of non-specific low back pain: a review of the litterature on classification systems relevant to physiotherapy. *Phys Ther Rev* 1999;**4**:265–81.
58. **Nwuga G**, Nwuga V. Relative therapeutic efficacy of the Williams and McKenzie protocols in back pain management. *Physiother Pract* 1985;**1**:99–105.
59. **McCarthy C**, Arnall F, Strimpakos N, *et al.* The biopsychosocial classification of non-specific low back pain: a systematic review. *Phys Therapy Rev* 2004;**9**:17–30.
60. **Petersen T**, Olsen S, Laslett M, *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.
61. **Kerns RD**, Turk DC, Rudy TE. The West Haven–Yale Multidimensional Pain Inventory (WHYMPI). *Pain* 1985;**23**:345–56.
62. **Turk DC**, Okifuji A, Sinclair JD, *et al.* Interdisciplinary treatment for fibromyalgia syndrome: clinical and statistical significance. *Arthritis Care Res* 1998;**11**:186–95.
63. **McKenzie R**. *The lumbar spine, mechanical diagnosis and treatment*. Waikanae: Spinal Publications Ltd, New Zealand, 1981.
64. Adams MA, Dolan P, Hutton WC, *et al*. Diural changes in the spinal mechanics and their clinical significance. *J Bone Joint Surg Br* 1990;**72B**:266–70.
65. **Borkan J**, Van Tulder M, Reis S, *et al.* Advances in the field of low back pain in primary care: a report from the fourth international forum. *Spine* 2002;**27**:E128–32.
66. **Waddell G**. *The back pain revolution*. Edinburgh: Churchill Livingstone, 2004.
67. **Dankaerts W**, O'Sullivan P, Burnett A, *et al.* Differences in sitting postures are associated with non-specific chronic low back pain disorders when patients are subclassified. *Spine* 2006;**31**:698–704.
68. **O'Sullivan P**. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005;**10**:242–55.

Integration of subclassification strategies in randomised controlled clinical trials evaluating manual therapy treatment and exercise therapy for non-specific chronic low back pain: a systematic review

K V Fersum, W Dankaerts, P B O'Sullivan, et al.

Br J Sports Med 2010 44: 1054-1062 originally published online December 8, 2009
doi: 10.1136/bjsm.2009.063289