# Clinimetrics corner: choosing appropriate study designs for particular questions about treatment subgroups

**Peter Kent[1,2], Mark Hancock[3], Ditte H. D. Petersen[4], Hanne L. Mjøsund[4]**

[1]Spine Centre of Southern Denmark, Middelfart Denmark, [2]Institute of Regional Health Services Research, University of Southern Denmark, Odense, Denmark, [3]Back Pain Research Group, Faculty of Health Sciences, The University of Sydney, Australia, [4]Department of Physiotherapy, University College Lillebælt, Odense, Denmark

**Background:** Many clinicians and researchers believe that there are subgroups of people with spinal pain who respond differently to treatment and have different prognoses. There has been considerable interest in this topic recently. However, problems occur when conclusions about subgroups are made that are inappropriate given the randomized controlled trial design used. The research design to choose, when developing a study protocol that investigates the effect of treatment subgroups, depends on the particular research question. Similarly, the inferences that can be drawn from an existing study will vary, depending on the design of the trial.

**Objectives:** This paper discusses the randomized controlled trial designs that are suitable to answer particular questions about treatment subgroups. It focuses on trial designs that are suitable to answer four questions: (1) 'Is the treatment effective in a pre-specified group of patients?'; (2) 'Are outcomes of treatment applied using a subgrouping clinical reasoning process, better than a control treatment?'; (3) 'Are the outcomes for a patient subgroup receiving a particular treatment (compared to a control treatment) better than for patients not in the subgroup who receive the same treatment?'; and (4) 'Are outcomes for a number of treatments better if those treatments are matched to patients in specific subgroups, than if the SAME treatments are randomly given to patients?'. Illustrative examples of these studies are provided.

**Conclusion:** If the clinical usefulness of targeting treatments to subgroups of people is to be determined, an important step is a shared understanding of what different RCT designs can tell us about subgroups.

**Keywords:** Classification, Methods, Randomized controlled trial, Subgroups

Many clinicians and researchers believe that there are subgroups of people with spinal pain who respond differently to treatment and have different prognoses.[1,2] If subgroups of patients who respond best to specific interventions could be identified, the potential exists to significantly improve patient outcomes and healthcare system efficiency. There has been a surge in interest in this topic with many recent studies and review articles.[3–14]

Randomized controlled trials (RCTs) are the gold standard for investigating treatment effects, due to the ability of this type of study to control for bias and produce precise measurements of treatment efficacy. Subgroup studies aiming to identify patient features associated with treatment effects also need to use an RCT design to provide strong evidence for treatment subgroups. When designs other than RCTs are used to identify subgroups, the clinical characteristics of some subgroups may include prognostic factors (indicative of likely outcomes regardless of treatment) rather than treatment effect modifiers (indicative of likely response to a specific treatment). Figure 1 demonstrates the importance of a control group (comparison treatment) in order to tease apart the treatment effect influence of subgroup membership (treatment effect modification) from the prognostic effect of subgroup membership.[9,15] Different RCT designs have been used to investigate treatment subgroups and each design is best suited to answering particular research questions about subgroups. It is important that the appropriate RCT design is used for a specific research question; otherwise, the validity of the conclusions will be compromised.

The subgroup research question being investigated will be influenced by the type of subgroup, the

Correspondence to: Peter Kent, Research Department, Spine Centre of Southern Denmark, Oestre Hougvej 55, 5500 Middelfart, Denmark. Email: peter.kent@slb.
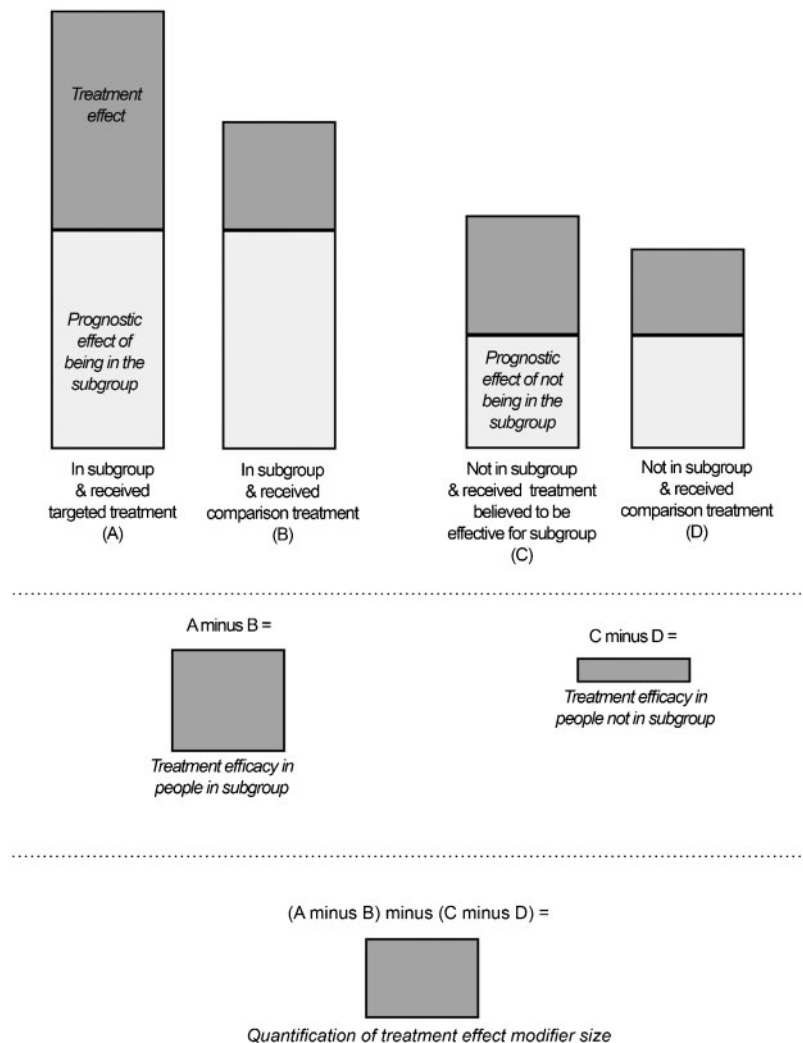
**Figure 1 Illustration of the importance of a comparison treatment in teasing apart the treatment modifier effects and prognostic effects of subgroup membership.**

treatment being investigated, and the comparison treatment. Studies investigating a dichotomous subgroup classification (such as belonging or not belonging to a subgroup) require a different design to those investigating a more complex reasoning process or algorithm that places patients into one of several subgroups (such as the treatment-based classification approach).[16] Identifying a subgroup of patients who respond to a specific treatment (for example, spinal manipulation) is different from investigating the efficacy of an algorithm that matches each patient to one of a range of different treatments (such as exercise, manipulation, or traction). Similarly, the comparison or control treatment is very important in subgroup studies. For example, comparing outcomes between a subgrouping approach (for example, the Sarhmann movement impairment-based approach)[17] and another quite different treatment approach (such as massage) answers a different question to comparing the subgrouping approach to the same interventions applied without the subgrouping system.

Problems occur when conclusions about subgroups are made that are inappropriate given the RCT design used. The aim of this paper is to discuss the

RCT designs that are suitable to answer particular questions about treatment subgroups.

## Different Designs for Different Questions
### Question 1: Is the treatment effective in a pre-specified group of patients?

This question is used when the research interest is only in how people included in the subgroup tested in the trial respond to treatment. That is, one is not concerned about how people who are not in this subgroup would respond to the same treatment. This would be the case if it were clearly unsafe and/or unethical to provide the treatment to other patients, or if there were a strong rationale as to why it would not be expected to be effective in other patients. Hypothetically, for example, a trial of lumbar discectomy for back-related leg pain compared to conservative care, might only include patients with MRI evidence of disc herniation. While it is theoretically possible that some people with back-related leg pain but no MRI evidence of herniation might be improved with discectomy, such people are not included in such a trial. While this trial will not tell us if discectomy is more effective for those with herniation than those without, it would be unethical
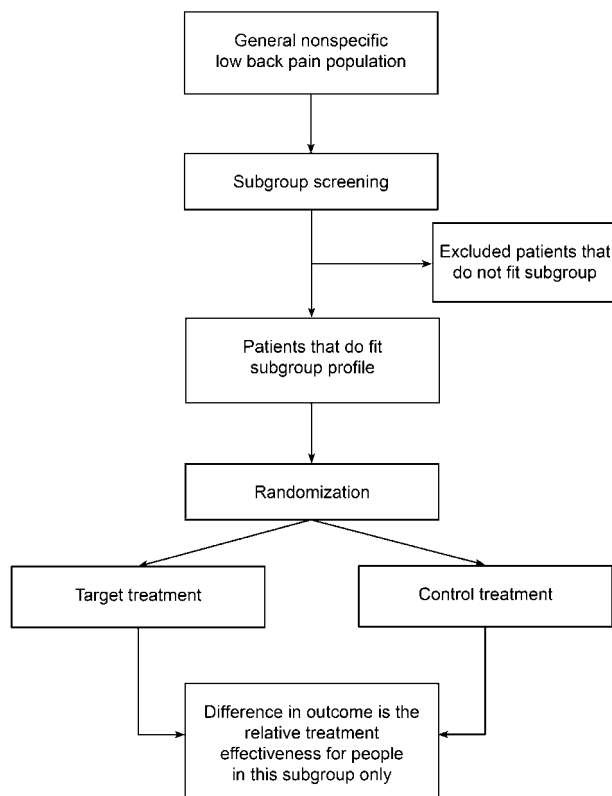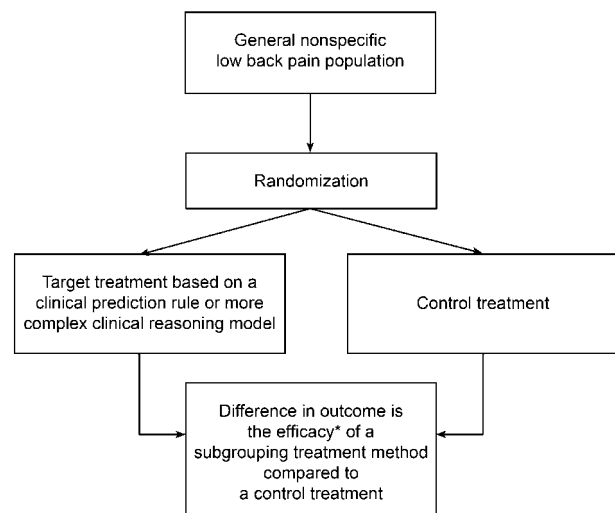
**Figure 2  Single-subgroup RCT design.**



* This efficacy is a combination of a treatment effect (how good is the treatment regardless of subgrouping) and a subgrouping effect (did targeting the treatment help?). Using this RCT design, it is not possible to differentiate between these two effects, unless the control treatment is the same as the target treatment, just without the subgroup clinical reasoning.

**Figure 3  Subgroup system RCT design.**

to enrol patients without any evidence of herniation. The key issue is that the results tell us about the efficacy of the treatment in the patients included in the trial (back-related leg pain and MRI evidence of herniation) but do not provide any evidence of superior efficacy in the patients included in the trial compared to those not included. To express this in another way, this design does not provide evidence of treatment effect modification. Furthermore, the results of the trial should only be generalized to patients similar to those included. We have called the design which addresses this question a 'single subgroup RCT design' (Fig. 2). Examples of studies using this approach include Browder *et al.*[18] (only included people with a directional preference), O'Sullivan *et al.*[19] (only included people with spondylolisthesis), and Cleland *et al.*[5] (only included people positive on the Flynn manipulation prediction rule).

The study by Cleland *et al.*[5] is an example where we believe that the study design is inappropriate for the aims and conclusions. The study randomized patients who were positive on the Flynn prediction rule to one of three manual therapy techniques. As only rule-positive people were included, it was not possible to determine treatment efficacy in people who were rule-negative. Therefore, there were insufficient data available to determine whether the relative efficacy of the treatments was the same regardless of prediction rule status and thereby test the validity of the rule. However, the authors conclude that the study provides evidence for the validity of the

prediction rule in identifying patients who respond to high velocity manipulation.

*Question 2: Are outcomes of treatment applied using a subgrouping clinical reasoning process, better than a control treatment?*

In this design, a subgrouping approach to treatment such as McKenzie therapy[20] is compared to another treatment (control treatment). This design is appropriate to determine if the overall treatment approach (including subgrouping) is superior to the control treatment but does not validate the subgrouping used, unless the treatments in the control group are the same but applied without the subgrouping. For example, comparing McKenzie therapy to a comparison treatment (such as anti-inflammatory medication) does not validate the subgroup classification, as any superior outcomes may simply be due to the superior efficacy of the exercises and other techniques used in the McKenzie approach, regardless of the subgroup in which a patient is classified. We have called the design which addresses this question a 'subgroup system RCT' (Fig. 3). Examples of this design include Chiradejnant *et al.*[21] (PA mobilisation), Descarreaux *et al.*[22] (muscle deficit assessment), Fritz *et al.*[23] (treatment-based classification), Geisser *et al.*[24] (manual medicine assessment), and Petersen *et al.*[25,26] (McKenzie assessment).

*Question 3: Are the outcomes for a patient subgroup receiving a particular treatment (compared to a control treatment) better than for patients not in the subgroup who receive the same treatment?*

To address this question, the design must randomize patients to either the targeted treatment group or a control treatment group, plus these two groups must also include patients whose clinical characteristics meet the subgroup criteria and those who do not.
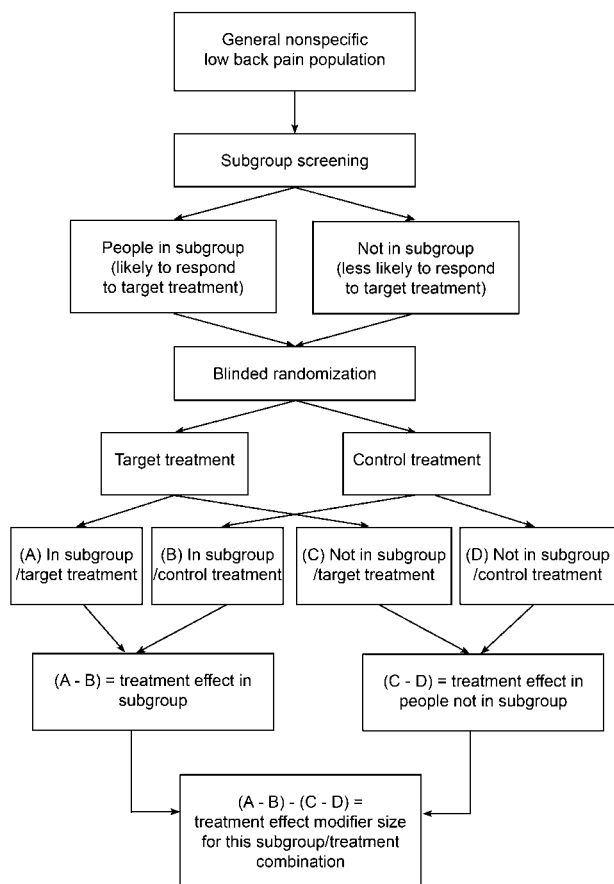
**Figure 4** Two-group plus subgroup covariate RCT design.

This design can be used whether the subgroup is defined by a single characteristic such as gender or when it is defined by a number of patient characteristics. This study design provides robust evidence that a patient subgroup really does exist whose members respond better to a particular treatment.[9] Using this design, the results of a statistical interaction test determine if those who are in the subgroup get more benefit from the target treatment (compared to the control treatment) than those who are not in the subgroup and the size of the benefit. We have called the study design that addresses this question, the 'two-group plus subgroup covariate RCT design' (Fig. 4). Examples of this study design are Childs *et al.*[4] and Hancock *et al.*[10] (both examining the Flynn manipulation prediction rule).

While this design provides robust evidence of a treatment subgroup (treatment modifier effect), the results are specific to the targeted treatment and the control treatment used. A subgroup who responds better to a particular treatment compared to placebo may not necessarily be the same as a subgroup who responds well to the same treatment compared to a different control treatment. For example, if a study demonstrates that reduced lumbar spine flexion identifies a subgroup who responds best to low velocity mobilization techniques compared to no treatment, it does not necessarily mean that reduced lumbar flexion necessarily identifies those who respond better to mobilization compared to a
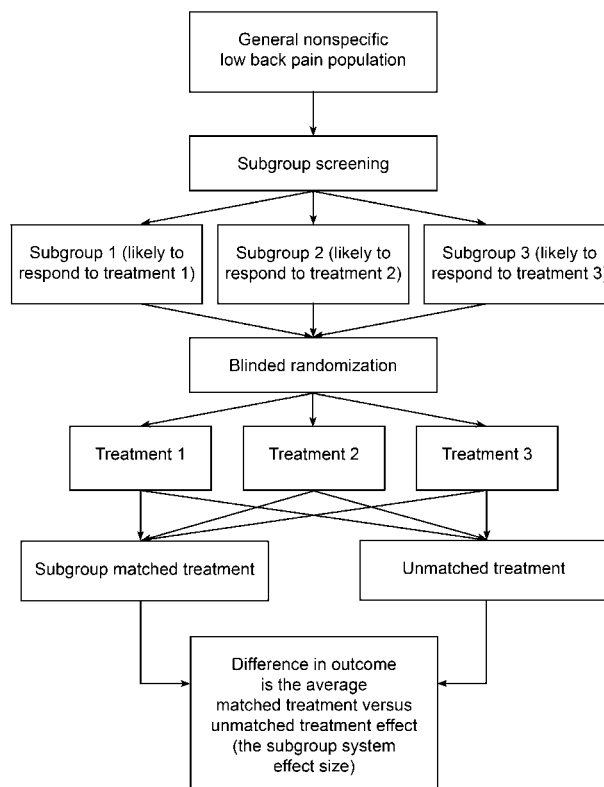
different control such as exercise therapy or high velocity manipulation. It is possible that reduced lumbar flexion also identifies patients who do well with these treatments (exercise therapy or high velocity manipulation) and as such is not useful to select between mobilization and these treatments. In other words, subgroups identified using this study design should only be considered to generalize to the treatment and control in that study, until shown to generalize to other comparisons. In addition, compared to a conventional two-group RCT, with this type of RCT, there is an increased likelihood that the results may be due to chance and so, there is also an increased need to validate the findings in other samples of patients.[9,27]

## Question 4: Are outcomes for a range of treatments better if those treatments are matched to patients in specific subgroups, than if the SAME treatments are randomly given to patients?

This question is used when the focus of the study is on the effectiveness of a subgroup or classification approach that includes a range of treatments. It is not focusing on the efficacy of the individual treatments but on whether outcomes are better if a subgroup system is used to target the treatments compared to when the same treatments are used but not matched to patients. We have called the design used to answer this question the 'multi-arm subgroup system RCT design' (Fig. 5).

An example of a study addressing this question is Brennan *et al.*,[28] which involved investigation of the efficacy of three individual treatments (manipulation,



**Figure 5** Multi-arm subgroup system RCT design.

stabilization exercises, and direction-specific exercises) to determine if these were better if targeted to subgroups of patients using a classification algorithm. Patients were classified as being suitable for one of the three treatments based on an initial evaluation and then randomized to one of the three treatments regardless of classification. The analysis compared the outcomes in patients who by chance received treatment matched to their classification, to those who received non-matched treatment. A second example of a study addressing this question is the study of directional preference exercises by Long *et al.*[12]

This research design can investigate a multi-subgroup system while controlling for prognostic and non-specific treatment effects. A potential limitation of this design is that it is usually not possible to isolate the efficacy of each subgroup/treatment component as only the combined targeted treatment effect is observed. For example, in the Brennan *et al.*'s study,[28] patients matched to the treatment had better outcomes than those not matched; however, it is not possible, based on the reported results to determine if the matching was effective for some of the three treatments (for example, manipulation and directional preference exercises) but not for others (such as stabilization exercises). The multi-arm subgroup system RCT design can be conceptualized as a number of two-group plus subgroup covariate RCTs run in parallel. If authors reported adequate information (outcomes for each treatment in those matched and not matched), each subgroup/treatment component could be analysed as a separate two-group plus subgroup covariate RCT, which would enable the relative efficacy of treatment matching for each subgroup to be investigated. However, each of these analyses would have to be adequately powered.[9,27]

A further caveat to use of the multi-arm subgroup system RCT design is that imperfections in the randomization process may result in an imbalance in the proportions of people allocated to the particular treatments in the matched and unmatched treatment groups. If this occurs, any observed targeted treatment effect may actually be due to differences in the efficacy of particular treatments, not due to the targeting of treatment. An additional prerequisite for this design is that all people in the population of interest fit one of the subgroups and that there are treatments available that people in each of these subgroups are likely to respond to.

## Summary

The research design to choose, when developing a study protocol that investigates the effect of treatment subgroups, depends on the particular research question (Table 1). Similarly, the inferences that can be drawn from an existing study will vary, depending on the design of the trial. This paper discussed the randomized controlled trial designs that are suitable to answer four questions: (1) 'Is the treatment effective in a pre-specified group of patients?'; (2) 'Are outcomes of treatment applied using a subgrouping clinical reasoning process, better than a control treatment?'; (3) 'Are the outcomes for a patient subgroup receiving a particular treatment (compared to a control treatment) better than for patients not in the subgroup who receive the same treatment?'; and (4) 'Are outcomes for a number of treatments better if those treatments are matched to patients in specific subgroups, than if the same treatments are randomly given to patients?'. Illustrative examples of these studies were provided.

Although the focus of the paper has been on studies investigating treatment subgroups in people with back pain, similar trial design and interpretation issues apply to studies of other health conditions. If we are to determine the clinical usefulness of targeting treatments to subgroups of people, a key step is a shared understanding of what different RCT designs can and cannot tell us about subgroups. Given the level of interest in the manual therapy community in subgrouping, arriving at such a shared understanding is important.

## Acknowledgements

**Table 1** Research questions about subgroups and the appropriate RCT design

| Question of interest | Appropriate design |
|---|---|
| (1) Is the treatment effective in a pre-specified group of patients? | Single subgroup RCT |
| (2) Are outcomes of treatment applied using a subgrouping clinical reasoning process, better than a control treatment? | Subgroup system RCT |
| (3) Are the outcomes for a patient subgroup receiving a particular treatment (compared to a control treatment) better than for patients not in the subgroup who receive the same treatment? | Two-group plus subgroup covariate RCT |
| (4) Are outcomes for a number of treatments better if those treatments are matched to patients in specific subgroups, than if the same treatments are randomly given to patients? | Multi-arm subgroup system RCT |

# References

1 Kent P, Keating JL, Buchbinder R. Searching for a conceptual framework for nonspecific low back pain. Man Ther 2009;14:387–96.

2 Kent PM, Keating J. Do primary-care clinicians think that non-specific low back pain is one condition? Spine (Phila Pa 1976) 2004;29:1022–31.

3 Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. Phys Ther 2006;86:122–31.

4 Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, *et al.* A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulations: a validation study. Ann Intern Med 2004;141:920–8.

5 Cleland JA, Fritz JM, Kulig K, Davenport TE, Eberhart S, Magel J, *et al.* Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: a randomized clinical trial. Spine (Phila Pa 1976) 2009;34:2720–9.

6 Cleland JA, Fritz JM, Whitman JM, Childs JD, Palmer JA. The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: a case series. J Orthop Sports Phys Ther 2006;36:198–9.

7 Fritz JM, Whitman JM, Childs JD. Lumbar spine segmental mobility assessment: an examination of validity for determining intervention strategies in patients with low back pain. Arch Phys Med Rehabil 2005;86:1745–52.

8 George SZ, Delitto A. Clinical examination variables discriminate among treatment-based classification groups: a study of construct validity in patients with acute low back pain. Phys Ther 2005;85:306–14.

9 Hancock M, Herbert R, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther 2009;89:698–704.

10 Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy. A randomised controlled trial. Eur Spine J 2008;17:936–43.

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

12 Long A, Donelson R, Fung T. Does it matter which exercise? A randomized control trial of exercise for low back pain. Spine (Phila Pa 1976) 2004;29:2593–602.

13 O'Sullivan PB, Beales DJ. Diagnosis and classification of pelvic girdle pain disorders, Part 2: Illustration of the utility of a classification system via case studies. Man Ther 2007;12:e1–12.

14 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.

15 Kamper SJ, Maher CG, Hancock MJ, Koes BW, Croft PR, Hay E. Treatment-based subgroups of low back pain: a guide to appraisal of research studies and a summary of current evidence. Best Pract Res Clin Rheumatol 2010;24:181–91.

16 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;75:470–89.

17 Sahrmann S. Diagnosis and treatment of movement impairment syndromes. St Louis, MO: Mosby; 2002.

18 Browder DA, Childs JD, Cleland JA, Fritz JM. Effectiveness of an extension-oriented treatment approach in a subgroup of subjects with low back pain: a randomized clinical trial. Phys Ther 2007;87:1608–18.

19 O'Sullivan PB, Twomey LT, Allison GT. Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis. Spine (Phila Pa 1976) 1997;22:2959–67.

20 McKenzie R, May S. Lumbar spine, mechanical diagnosis and therapy. 2nd ed. Waikanae: Spinal Publications Ltd; 2003.

21 Chiradejnant A, Latimer J, Maher CG, Stepkovitch N. Does the choice of spinal level treated during posteroanterior (PA) mobilisation affect treatment outcome? Physiother Theory Pract 2002;18:165–74.

22 Descarreaux M, Normand MC, Laurencelle L, Dugas C. Evaluation of a specific home exercise program for low back pain. J Manipulative Physiol Ther 2002;25:497–503.

23 Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. Spine (Phila Pa 1976) 2003;28:1363–72.

24 Geisser ME, Wiggert EA, Haig AJ, Colwell MO. A randomized, controlled trial of manual therapy and specific adjuvant exercise for chronic low back pain. Clin J Pain 2005;21:463–70.

25 Petersen T, Kryger P, Ekdahl C, Olsen S, Jacobsen S. 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 randomized controlled trial. Spine (Phila Pa 1976) 2002;27:1702–9.

26 Petersen T, Larsen K, Jacobsen S. One-year follow-up comparison of the effectiveness of McKenzie treatment and strengthening training for patients with chronic low back pain: outcome and prognostic factors. Spine (Phila Pa 1976) 2007;32:2948–56.

27 Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229–36.

28 Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. Spine (Phila Pa 1976) 2006;31:623–31.