Journal of Clinical Epidemiology 79 (2016) 3–9

Journal of Clinical Epidemiology

REVIEWS

A systematic review reveals that the credibility of subgroup claims in low back pain trials was low

Bruno T. Saragiotto[a,*], Chris G. Maher[a], Anne M. Moseley[a], Tie P. Yamato[a], Bart W. Koes[b], Xin Sun[c,d], Mark J. Hancock[e]

[a]*Musculoskeletal Division, The George Institute for Global Health, Sydney Medical School, The University of Sydney, Level 3/50 Bridge Street, 2000, Sydney, NSW, Australia*

[b]*Department of General Practice, Erasmus MC, South Holland, Wytemaweg 80, 3015, CN, Rotterdam, The Netherlands*

[c]*Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Section 3, Ren Min Nan Lu. Chengdu, Sichuan 610041, P.R. China*

[d]*Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, 1280 Main Street West, L8S4L8, Hamilton, Ontario, Canada*

[e]*Faculty of Medicine and Health Sciences, Macquarie University, Balaclava Road, North Ryde, Sydney, New South Wales 2109, Australia*

Accepted 3 June 2016; Published online 10 June 2016

Abstract

Objectives: To assess the credibility of subgroup claims in back pain randomized controlled trials.

Study Design and Setting: A sample of reports of back pain trials from 2000 to 2015 that provided a subgroup claim were included ($n = 38$). Two reviewers independently assessed risk of bias and credibility of subgroup claims as well as the strength of the author's claim. The credibility of subgroup claims was assessed using a 10-criteria tool, and strength of the subgroup claims was assessed based on seven criteria to categorize claims into a reasonably strong claim of a definitive subgroup effect or a more cautious claim of a possible effect.

Results: A total of 91 claims of a subgroup effect were reported in the 38 included trials, of which 28 were considered strong claims of a definitive effect, and 63 were cautious claims of a possible effect. None of the subgroup claims met all 10 credibility criteria, and only 24% (22 claims) satisfied at least five criteria. The only criteria satisfied by more than 50% of the claims were if the subgroup variable was a characteristic measured at baseline, and whether the test of interaction was significant. All other criteria were satisfied by less than 30% of the claims. There was no association between the credibility of subgroup claims and the journal impact factor, risk of bias, sample size, or year of publication.

Conclusion: The credibility of subgroup claims in back pain trials is usually low, irrespective of the strength of the authors' claim.

Keywords: Back pain; Research design; Effect modifier; Subgroup analysis; Methods; Treatment outcome

1. Introduction

Low back pain (LBP) is the leading cause of years lived with disability worldwide [1]. The point prevalence of LBP is estimated to be 9.4% [2], while the lifetime prevalence can reach up to 39% [3]. The minority of people with LBP seen in primary care receive a specific diagnosis for the cause of their LBP (e.g., infection, tumor, ankylosing spondylitis, fracture, radicular syndrome, or cauda equina syndrome). For the great majority of people (about 90%), the source of pain cannot be determined with conventional tests and they are classified as having nonspecific LBP. According to clinical practice guidelines, these people with nonspecific LBP should be managed with generic treatments such as analgesic medicines, physical therapies (e.g., exercise, spinal manipulation), or psychological therapies [4,5].

It has been suggested that the large group of people with nonspecific LBP could be divided into subgroups of people who will respond better to one treatment than another [6,7]. This approach, based on subgroups, offers the possibility of a larger treatment effect than applying generic treatments to people with nonspecific LBP [8,9]. The identification of subgroups has been proposed as an important research priority internationally [10,11], and subgroup analyses have been included in several randomized controlled trials in the field of LBP [8,12–15]. However, methodological limitations

* Corresponding author. Level 3/ 50 Bridge Street, Sydney, New South Wales 2000, Australia.Tel.: +61-2-8052-4300; fax: +61-2-8052-4301.

E-mail address: bsaragiotto@georgeinstitute.org.au (B.T. Saragiotto).

http://dx.doi.org/10.1016/j.jclinepi.2016.06.003

What is new?

Key findings

- The credibility of subgroup claims in back pain trials is usually low, with the potential to mislead clinicians. The strength of the author's claim of a subgroup effect is often overstated and not tempered by the credibility of subgroup claims.

What this adds to what was known?

- This is the first study to investigate the credibility of subgroup claims in low back pain using standardized criteria for assessing credibility and a comprehensive search for the identification of studies.

What is the implication and what should change now?

- The credibility criteria used in this review can help guide researchers to improve the conduct and reporting of subgroup analyses. Research authors should include an overview of the credibility of their subgroup analysis to avoid overstating the strength of subgroup effects.

such as failing to prespecify the hypothesis of the subgroup effect, performing a large number of post hoc subgroup analyses, or statistical analysis performed inappropriately make the findings susceptible to several biases [16,17].

The need for standards for the interpretation of subgroup analyses is clear [18]. To guide interpretation of trial reports claiming subgroup effects, explicit criteria have been developed to judge their credibility: i.e. whether a reported difference in a treatment effect between subgroups is likely to be real or not [19,20]. Sun et al. [16] investigated the credibility of subgroup claims in randomized controlled trials of medical interventions published in 2007. The authors found that the credibility of subgroup effects in most trials was usually low, with insufficient evidence to support the subgroup claims, statistical analysis performed inappropriately, and current evidence contradictory with the author's claims (e.g., inconsistency of effect across external studies). Mistaken claims of subgroup effects may result in people being denied a beneficial treatment or even receiving a potentially harmful or ineffective treatment.

While there are some review articles considering credibility of subgroup analyses based on clinical prediction rules [21,22], this represents only a subset of LBP trials that undertake subgroup analyses. Accordingly, investigating the credibility of treatment-based subgroup analyses in LBP trials would represent an important advance in knowledge [9]. The aim of this study was to assess the credibility of subgroup claims in reports of randomized controlled trials evaluating treatments for nonspecific LBP. We also investigated what importance authors placed on the subgroups they identify and relate the importance placed on the findings to the credibility of the subgroup claims. We hypothesized that the claims of subgroup effects in LBP trials would be mostly of low credibility.

2. Methods

This review was registered in the international prospective register of systematic reviews (PROSPERO 2014:CRD42014013063).

2.1. Types of studies

Published reports of randomized controlled trials evaluating treatment of LBP that make a subgroup claim for at least one outcome were included. Using the definition from a previous study [23], we considered a subgroup claim to have been made when the investigators stated in the abstract, results, or discussion that the effect of intervention differed, or may have differed, according to the status of a subgroup variable. We only included trial reports that claimed a subgroup effect; trials that included a subgroup analysis but did not claim an effect were not included.

We defined a subgroup analysis as "a statistical analysis that explores whether effects of the intervention (i.e., experimental vs. control) differ according to status of a subgroup variable." This includes post hoc or secondary analysis of the main result or former trial. In addition, we defined a subgroup effect as "a difference in the magnitude of a treatment effect across subgroups of a study population" [23].

2.2. Types of participants

Inclusion criteria:

- Use of true randomization (trials that use methods that are intended to be random, e.g., alternation, will be excluded).
- Trials evaluating participants with acute, subacute, chronic, or recurrent LBP or any combination.
- Trials that recruited participants from primary, secondary, or tertiary care, either seeking care for LBP or recruited from the community.
- Trials that reported a claim of a subgroup effect.

Exclusion criteria

- Studies evaluating specific forms of back pain (e.g., cancer, fracture, cauda equina syndrome, and inflammatory diseases).

2.3. Types of interventions

We considered any type of intervention used for treating LBP (including surgical, pharmacologic, psychological,

physical, and alternative medicine). We also considered any type of treatment comparison in the included studies.

2.4. Types of outcomes measures

We considered any type of outcome measure that could be considered patient centered (e.g., pain, disability, quality of life).

2.5. Search methods for identification of studies

Searches were conducted on the Cochrane Central Register of Controlled Trials (CENTRAL), EMBASE, and MEDLINE databases from 2000 to April 2015. We only included published full-text articles, so conference abstracts, theses, and other unpublished reports were excluded. Trial reports written in English, Portuguese, Spanish, Dutch, and Chinese were able to be included as the review authors could read these languages. Reports written in other languages were excluded.

After the screening of titles and abstracts, the full text of potentially eligible studies was retrieved from three random samples of the original search, each of 250 records, plus all studies from an updated search. The records were imported into Microsoft Excel, and then, the random number function was used to identify the three random samples. Pairs of reviewers independently screened the full text of the samples of potentially eligible studies for studies which reported a subgroup effect and, where a subgroup claim was reported, and recorded the variables used to form the subgroups (e.g., age, gender, pain level). Disagreements between the reviewers were resolved through discussion or by arbitration of a third reviewer when consensus could not be reached.

2.6. Data extraction

For the data extraction, we used additional sources referenced in the included study (i.e., trial register, published protocol, online supplements, and primary and secondary trial reports). These documents were identified and obtained by a single reviewer. The following data were extracted from the included studies by a single reviewer:

1. Bibliometric data (authors, year of publication, language)
2. Sample size (number of individuals randomized)
3. Impact factor of the journal for 2014 (using the ISI Journal Citation Report)
4. Characteristics of the participants (gender, age, and duration of symptoms)
5. Broad description of the interventions and cointerventions
6. Method of identifying subgroups and subgroups identified (number of subgroups, *P* values, tests of interaction, studies cited for replication of subgroup effect).

The credibility of subgroup claims was assessed using the 10 criteria developed by Sun et al. [16]. These criteria have been widely used over the last 20 years [16,19,20,24] and are recommended for assessing how much confidence to place in the results of subgroup analyses [25]. To achieve a more consistent assessment, the coding options were revised and approved by the original author of the criteria. The coding options are listed in Appendix A at www.jclinepi.com. Assessment of credibility was done separately for each subgroup claim in each included study. The strength of the claim of a subgroup effect (i.e., the author's conviction) was determined based on methods used in previous studies [16,23] with the subgroup claim categorized as (1) reasonably strong claim of a definitive subgroup effect (the authors convey a conviction that the subgroup effect truly exists) or (2) cautious claim of a possible effect (the authors suggest a subgroup effect but convey uncertainty about whether such an effect exists). The seven criteria for the classification of strength of claim are in Appendix B at www.jclinepi.com. If the subgroup claim had at least four criteria classified as a strong claim, this subgroup claim was classified as a reasonably strong claim of a definitive effect. Classification of the strength of the subgroup effect claim was done separately for each subgroup in each included study.

Pairs of reviewers performed the assessment of credibility of subgroup effect and strength of claim. Any disagreements were resolved by discussion or arbitration from a third reviewer.

2.7. Assessment of risk of bias

Risk of bias was assessed using the Physiotherapy Evidence Database (PEDro) scale. The PEDro scale is composed by 10 items: random allocation, concealed allocation, similarity at baseline, subject blinding, therapist blinding, assessor blinding, $>85\%$ follow-up for at least one key outcome, intention-to treat analysis, between-group statistical comparison for at least one key outcome, and point and variability measures for at least one key outcome. Items are scored as either yes or no, and a score out of 10 is obtained by summation. The PEDro scale has been found to have acceptable reliability [26] and validity [27,28]. If the included study was indexed on PEDro (www.pedro.org.au), the total PEDro scores were downloaded from PEDro. If the included study was not indexed on PEDro, two independent reviewers (B.T.S. and T.P.Y.) performed the risk of bias assessment and possible disagreements between reviewers were resolved by discussion or arbitration by a third reviewer (A.M.M.) when consensus could not be reached.

2.8. Data analysis

The interrater agreement for the assessment of credibility of the subgroup claims was calculated using the

prevalence and bias adjusted kappa (PABAK) for each criteria. Interpretation of the PABAK coefficient has been described as: <0, poor; 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial, and 0.81–1.00, almost perfect [29].

We calculated the proportion of claims meeting each of the credibility criteria, and the number of credibility criteria met by each claim. Chi-square test or Fisher's exact test was used to compare the proportions meeting each of the criteria, and the number of criteria met by each subgroup. We also examined whether stronger claims met more credibility criteria using the test for trend. In addition, we conducted a multiple regression analysis to investigate the association between the credibility of subgroup claims (scores of the credibility items) and (1) risk of bias (total PEDro score); (2) publication year; (3) sample size; and (4) impact factor of the journal. The journal impact factor was transformed to a five-point ordinal scale with journals without an impact given a 0 score and those with an impact factor stratified by quartiles (i.e., no impact factor available, 1st quartile, 2nd quartile, 3rd quartile, and 4th quartile).

3. Results

The searches yielded 14,889 records, of which 2,451 were considered potentially eligible (i.e., randomized controlled trials in LBP). Based on a previous study [23], we anticipated in our protocol that about 20% of LBP trials would report a subgroup effect, and therefore, two random samples of 250 articles would be needed to reach our target of 100 studies. However, we screened three random samples of 250 potentially eligible records (750 records) plus 636 records from the updated search 1 year after the initial search, to accrue 66 trials that reported a subgroup analysis. The decision to update the initial search was made based on the assumption that the prevalence of subgroup analysis in trials of LBP has increased over recent years; thus, screening the update would improve the number of trials included in this study. Of the 66 trials that reported a subgroup analysis, 38 trials claimed a subgroup effect and were included in this study (Fig. 1). The characteristics of the included studies are presented in Table 1.

Among the 38 included studies, 14 made 1 claim of a subgroup effect, 10 studies made 2 claims, 7 studies made 3 claims, 3 studies made 4 claims, and 4 studies made 5 or more claims. Thus, a total of 91 claims of a subgroup effect were reported in the included studies. All claims of subgroup effects are listed in Appendix C at www.jclinepi.com. The interrater agreement for the assessment of the credibility of the subgroup claims ranged from 80% to 99%, and the PABAK values ranged from 0.61 to 0.98, representing moderate to almost perfect agreement.

Twenty-eight subgroup claims were considered strong claims of a definitive effect, and 63 were cautious claims of a possible effect (Table 2). None of the subgroup claims met all 10 credibility criteria, and only 22 (24%) claims satisfied at least five criteria. In the 91 claims, the only criteria satisfied by more than 50% of the included studies were if the subgroup variable was a characteristic measured at baseline, and whether the test of interaction was significant. Two criteria (indirect evidence to support the effect and if the subgroup effect was consistent across related outcomes) were significantly more likely satisfied in reports making strong claims compared to those with cautious claims (Table 2). All other criteria were satisfied by less than 30% of the included studies, irrespective of the strength of the claim. None of the studies correctly prespecified the subgroup hypotheses, and only six claims (7%) met the criteria for prespecifying the subgroup hypotheses. The median (interquartile range) number of criteria

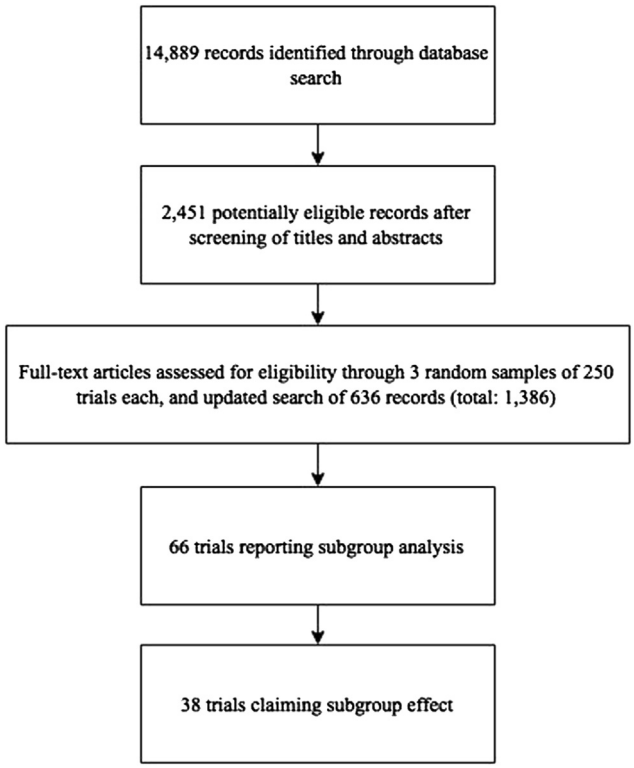

Fig. 1. Flowchart of the inclusion process of the review.

Table 1. Characteristics of included studies ($n = 38$)

Variable	Median (interquartile range) or frequency (%)
Sample size, median (IQR)	165.5 (113.5–270.8)
Age, median (IQR)	42.8 (40.8–47.9)
Total PEDro score, median (IQR)	7.0 (5.0–7.0)
Journal impact factor, median (IQR)	2.30 (1.76–2.78)
Duration of symptoms, *n* (%)	
Acute	2 (5)
Subacute	6 (16)
Chronic	30 (79)
Funding (yes), *n* (%)	29 (76)

Abbreviation: IQR, interquartile range.

Table 2. Number (percentage) of claims meeting each quality criteria

Criteria	Strong claim of a definitive effect ($n = 28$)		Cautious claim of possible effect ($n = 63$)		*P*-value
	Yes	No	Yes	No	
1. Is the subgroup variable a characteristic measured at baseline?	27 (96.4)	1 (3.6)	60 (95.2)	3 (4.8)	0.798[N.S.]
2. Was the subgroup variable a stratification factor at randomization?	1 (3.6)	27 (96.4)	4 (6.3)	59 (93.7)	0.591[N.S.]
3. Was the hypothesis specified a priori?	0 (0)	28 (100)	6 (9.5)	57 (90.5)	0.091[N.S.]
4. Was the subgroup analysis one of small number of subgroup hypotheses tested (≤5)?	8 (28.6)	20 (71.4)	18 (28.6)	45 (71.4)	1.000[N.S.]
5. Was the test of interaction significant (interaction $P < 0.05$)?	25 (89.3)	3 (10.7)	48 (76.2)	15 (23.8)	0.148[N.S.]
6. Was the significant interaction effect independent, if there were multiple significant interactions?	1 (3.6)	27 (96.4)	2 (3.2)	61 (96.8)	0.922[N.S.]
7. Was the direction of subgroup effect correctly prespecified?	0 (0)	28 (100)	0 (0)	63 (100)	—
8. Was the subgroup effect consistent with evidence from previous studies?	6 (21.4)	22 (78.6)	12 (19)	51 (81)	0.792[N.S.]
9. Was the subgroup effect consistent across related outcomes?	14 (50)	14 (50)	12 (19)	51 (81)	0.003*
10. Was there indirect evidence to support the apparent subgroup effect (biological rationale, laboratory tests, animal studies)?	21 (75)	7 (25)	28 (44.4)	35 (55.6)	0.007*

Abbreviation: N.S., not statistically significant.
*$P < 0.01$.

satisfied by the studies making claims was 3 (2 to 4). In addition, there was no significant association between the credibility of a subgroup claim with the journal impact factor of the study, risk of bias, sample size, and year of publication (Table 3).

4. Discussion

The credibility of subgroup claims in LBP trials was low. Of the 91 claims of a subgroup effect identified, 28 were strong claims yet only 10 of the 28 claims were able to satisfy at least half of the credibility criteria and none satisfied all criteria. Having the subgroup variable measured at baseline and the use of an interaction test were the only criteria satisfied by more than 50% of the claims. Only two subgroup credibility criteria (subgroup effect consistent across related outcomes and support the subgroup claim with indirect evidence) were more commonly positive in studies making a strong claim, and strong claims seem more likely to find a subgroup effect consistent across related outcomes and support the subgroup claim with indirect evidence. All the included studies failed to prespecify the correct direction of the subgroup hypotheses, and the hypothesis was prespecified for only six claims. Furthermore, the credibility of the subgroup claims does not appear to be associated with other study factors, such as journal impact factor, risk of bias, sample size, or year of publication.

The strengths of this review include the use of standardized criteria for assessing credibility of subgroup claims and a comprehensive search for the identification of studies. The credibility criteria used in this review were first described in 1992 [19], then restructured to address more domains in recent publications [16,20]. The low number of subgroup claims identified is the main limitation of this study. Although we identified a large number of LBP trials, only a small number of these trials reported a subgroup analysis and consequently were eligible to make a subgroup claim. It seems that while endorsement of treatment subgroups is common, there is only limited evidence in the literature. In contrast to our prediction that about 20% of LBP trials would report a subgroup analysis, only 3% reported subgroup analyses. Because the low number of trials identified in the three random samples of potentially eligible trials from our first search, we decided to update our literature search to identify additional eligible studies. This decision, however, was not specified in our protocol and needs to be acknowledged as a limitation of this study.

This low rating of credibility of subgroup claims in LBP trials is consistent with pharmacologic studies. Sun et al. [16] reported low credibility of subgroup claims in pharmacokinetic studies, in which most trials failed to prespecify the hypotheses or present significant interaction tests. A recent review [30] investigated the quality of subgroup analyses in therapist-delivered interventions for LBP and concluded that subgroup analyses in LBP are typically

Table 3. Regression analysis on factors associated with the credibility of subgroup claims

Variables	Univariate analysis		Multivariate analysis	
	Regression coefficient (95% CI)	*P*-value	Regression coefficient (95% CI)	*P*-value
Journal impact factor	0.25 (−0.07, 0.57)	0.127*	0.23 (−0.10, 0.56)	0.169[N.S.]
Methodological quality	−0.11 (−0.29, 0.08)	0.248*	−0.09 (−0.27, 0.10)	0.363[N.S.]
Sample size	−0.0002 (−0.0006, 0.0002)	0.270[N.S.]	—	—
Year of publication	0.04 (−0.10, 0.17)	0.595[N.S.]	—	—

Abbreviations: N.S., not statistically significant; CI, confidence interval.
*$P > 0.25$.

underpowered with poor quality of reporting. Both reviews reported the failure to specify the subgroup hypotheses a prior as a common problem in trials, which is also consistent with our findings. Furthermore, Chan et al. [31] found that some subgroup analyses reported to be prespecified were not described in the trial protocols and registration documents; thus, subgroup analyses with clear evidence that they were prespecified appear to be even more rare.

In this study, we did not find any association between the credibility of subgroup claims with journal impact factor, risk of bias, sample size, and year of publication. It seems that the credibility of subgroup claims is not influenced by these variables. An alternate variable which may be related to the credibility of subgroup claims is trial funding, with industry-funded trials being more likely to report subgroup analyses and less likely to prespecify subgroup hypotheses [32]. Although we did not investigate the influence of funding in this review, most of the included studies were funded (76%).

Subgroup analyses are potentially important and, when performed correctly, could help individualize treatment for heterogeneous conditions like nonspecific LBP. However, the great majority of existing subgroup analyses in the literature have the potential to mislead clinicians. If authors do not provide an overview of the credibility of their own analyses or alert clinicians on the exploratory nature of the data, subgroup analyses could result in clinicians changing their practice based on spurious data and the possibility of patients receiving potentially harmful treatments. The credibility criteria used in this review can help guide researchers to improve the conduct and reporting of better subgroup analyses. Provision of a completed checklist as a table or electronic supplement by trials undertaking subgroup analyses would help clinicians and researchers to judge whether the findings should be used for guiding clinical practice or as hypothesis generation, making the results from subgroup analyses less likely to be misinterpreted.

5. Conclusion

The credibility of subgroup claims in back pain trials is usually low, irrespective of the strength of the authors' claim. Future studies aiming to perform subgroup analyses should follow specific guidelines to improve the credibility of their findings. Clinicians should be aware of the problem of spurious subgroup claims for clinical practice.

Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jclinepi.2016.06.003.

References

[1] Global Burden of Disease Study. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014;386:743–800.

[2] Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73:968–74.

[3] Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012;64:2028–37.

[4] Chou R, Qaseem A, Snow V, Casey D, Cross JT Jr, Shekelle P, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007;147:478–91.

[5] van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006;15(Suppl 2):S169–91.

[6] Kent P, Keating J. Do primary-care clinicians think that nonspecific low back pain is one condition? Spine (Phila Pa 1976) 2004;29:1022–31.

[7] Kent P, Keating JL. Classification in nonspecific low back pain: what methods do primary care clinicians currently use? Spine (Phila Pa 1976) 2005;30:1433–40.

[8] Fritz JM, Lindsay W, Matheson JW, Brennan GP, Hunter SJ, Moffit SD, et al. Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. Spine (Phila Pa 1976) 2007;32:E793–800.

[9] 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.

[10] Borkan JM, Koes B, Reis S, Cherkin DC. A report from the second international forum for primary care research on low back pain. Re-examining priorities. Spine (Phila Pa 1976) 1998;23:1992–6.

[11] Guccione A, Goldstein M, Elliott S. Clinical research agenda for physical therapy. Phys Ther 2000;80:499–513.

[12] 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.

[13] Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine (Phila Pa 1976) 2002;27:2835–43.

[14] 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 manipulation: a validation study. Ann Intern Med 2004;141:920–8.

[15] 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(6):623–31.

[16] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ 2012;344:e1553.

[17] van Klaveren D, Vergouwe Y, Farooq V, Serruys PW, Steyerberg EW. Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions. J Clin Epidemiol 2015;68:1366–74.

[18] Hasford J, Bramlage P, Koch G, Lehmacher W, Einhaupl K, Rothwell PM. Inconsistent trial assessments by the National Institute for Health and Clinical Excellence and IQWiG: standards for the performance and interpretation of subgroup analyses are needed. J Clin Epidemiol 2010;63:1298–304.

[19] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78–84.

[20] Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 2010;340:c117.
[21] Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Phys Ther 2010;90:843−54.
[22] Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther 2009;89:698−704.
[23] Sun X, Briel M, Busse JW, Akl EA, You JJ, Mejza F, et al. Subgroup Analysis of Trials Is Rarely Easy (SATIRE): a study protocol for a systematic review to characterize the analysis, reporting, and claim of subgroup effects in randomized trials. Trials 2009;10:101.
[24] Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. JAMA 2014;311:405−11.
[25] Oxman AD. Subgroup analyses. BMJ 2012;344:e2022.
[26] Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83:713−21.
[27] de Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. Aust J Physiother 2009;55:129−33.
[28] Macedo LG, Elkins MR, Maher CG, Moseley AM, Herbert RD, Sherrington C. There was evidence of convergent and construct validity of physiotherapy evidence database quality scale for physiotherapy trials. J Clin Epidemiol 2010;63:920−5.
[29] Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159−74.
[30] Mistry D, Patel S, Hee SW, Stallard N, Underwood M. Evaluating the quality of subgroup analyses in randomized controlled trials of therapist-delivered interventions for nonspecific low back pain: a systematic review. Spine (Phila Pa 1976) 2014;39:618−29.
[31] Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ 2008;337:a2299.
[32] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. BMJ 2011;342:d1569.