

Trends over time in the size and quality of randomised controlled trials of interventions for chronic low-back pain

Nicholas Henschke · Ton Kuijpers · Sidney M. Rubinstein ·
Marienke van Middelkoop · Raymond Ostelo · Arianne Verhagen ·
Bart W. Koes · Maurits W. van Tulder

Received: 18 January 2011 / Revised: 10 July 2011 / Accepted: 6 September 2011 / Published online: 27 October 2011
© Springer-Verlag 2011

Abstract

Purpose Previous reviews of randomised controlled trials (RCTs) for low-back pain (LBP) have failed to identify any positive trend in study quality with more recent years of publication. This study aimed to identify and describe trends over time in the study design characteristics and risk of bias in chronic LBP trials performed over the past 30 years.

Methods One fifty-seven randomised trials of interventions for chronic LBP were extracted from recently published systematic reviews. The reviews included RCTs on physical and rehabilitation interventions, injection therapy and denervation procedures, complementary and alternative therapies and pharmacological interventions for chronic LBP. Study level data were extracted and analysed for trends associated with year of publication.

Results Overall, the mean sample size in the RCTs was 141 (median 70; range 17–3093). There was a slight increase in the median number of risk of bias criteria fulfilled from trials published prior to 1995 to those published after 1996. The analysis showed that in more recent years RCTs of medical interventions were more likely to be successfully blinded than RCTs of non-medical interventions.

Conclusions The continuing uncertainty regarding the efficacy of many interventions for chronic LBP again stresses the need for large RCTs with low risk of bias. Further research is needed into specific risks of bias within the RCTs for chronic LBP and the effect they have on the plausibility of the results.

Keywords Back pain · Bias · Methodology · Randomised trials

N. Henschke (✉)

The George Institute for Global Health, Missenden Rd,
PO Box M201, Sydney, NSW 2050, Australia
e-mail: nhenschke@georgeinstitute.org.au

N. Henschke · S. M. Rubinstein · R. Ostelo · M. W. van Tulder
EMGO+ Institute for Health and Care Research,
Amsterdam, The Netherlands

T. Kuijpers
Dutch Institute for Healthcare Improvement CBO,
Utrecht, The Netherlands

M. van Middelkoop · A. Verhagen · B. W. Koes
Department of General Practice, Erasmus MC,
University Medical Centre, Rotterdam, The Netherlands

R. Ostelo · M. W. van Tulder
Department of Health Sciences, Faculty of Earth and Life
Sciences, VU University, Amsterdam, The Netherlands

Introduction

Low-back pain (LBP) is one of the most prevalent musculoskeletal conditions placing an enormous burden on societies and healthcare systems around the world [1]. Chronic LBP (which is the pain lasting for longer than 3 months) is responsible for a disproportionate amount of the economic and occupational costs of LBP. Moreover, its prevalence appears to be increasing [2]. In order to manage the increasing burden of chronic LBP, clinical guidelines have been developed to provide evidence-based recommendations to guide practice [3]. These guidelines synthesise the available evidence around management strategies for chronic LBP, but there remains uncertainty regarding the most effective form of therapy. This continuing uncertainty is often considered to be due to methodological

shortcomings or a lack of adequate reporting in randomised controlled trials (RCTs) which evaluate interventions for LBP [4].

The number of RCTs and systematic reviews on interventions for LBP has increased steadily over the past decades [5]; however, the quantity of studies alone is not sufficient for a valid evaluation of the efficacy of the therapeutic interventions. In recent years, several initiatives have been taken by the scientific community to improve the reporting of RCTs and to raise awareness of potential sources of bias within the RCTs. The impact of scientific journals adopting standards of reporting, such as the CONSORT statement [6, 7] has been reported as a positive improvement towards the reduction of the risks of bias [8]. The Cochrane Back Review Group (CBRG) provides method guidelines for conducting systematic reviews in the field of LBP which recommend a criteria list to evaluate the risks of bias within a RCT [9, 10]. Using a similar criteria list, Koes et al. [5, 11] reported that while high-quality RCTs are available on conservative interventions for LBP, they did not observe any trend toward higher methodological quality with more recent years of publication. In light of the increasing number of RCTs on interventions for chronic LBP and these continuing initiatives to improve the quality of clinical research in general, it could be expected that the methodological quality of RCTs in the field of chronic LBP may have improved during the recent years.

The objective of this study is to identify and describe trends over time in the study design characteristics and risk of bias in more than 150 RCTs of interventions for chronic LBP performed over the past 30 years.

Materials and methods

Four systematic reviews, commissioned by the Dutch Health Care Insurance Council (CVZ), were performed to evaluate current evidence for physical and rehabilitation interventions [12], complementary and alternative therapies [13], pharmacological interventions [14], and injection and denervation procedures [15] in patients with chronic LBP. The current study is based on the trials that were included in this series of systematic reviews. The reviews included only RCTs which evaluated therapeutic interventions in adult subjects (aged 18 years or older) with chronic (>12 weeks duration) non-specific LBP. All of the reviews followed similar methodology which was based on the recommendations of the CBRG [10] and which is described in more detail in the published reviews. Literature searching was performed on electronic databases such as MEDLINE and EMBASE, as well as by identifying eligible RCTs from previous systematic reviews within the

Cochrane Library. The risk of bias in all eligible trials was assessed independently by two authors using an 11 item criteria list [9] with disagreements resolved by discussion or consultation with a third author.

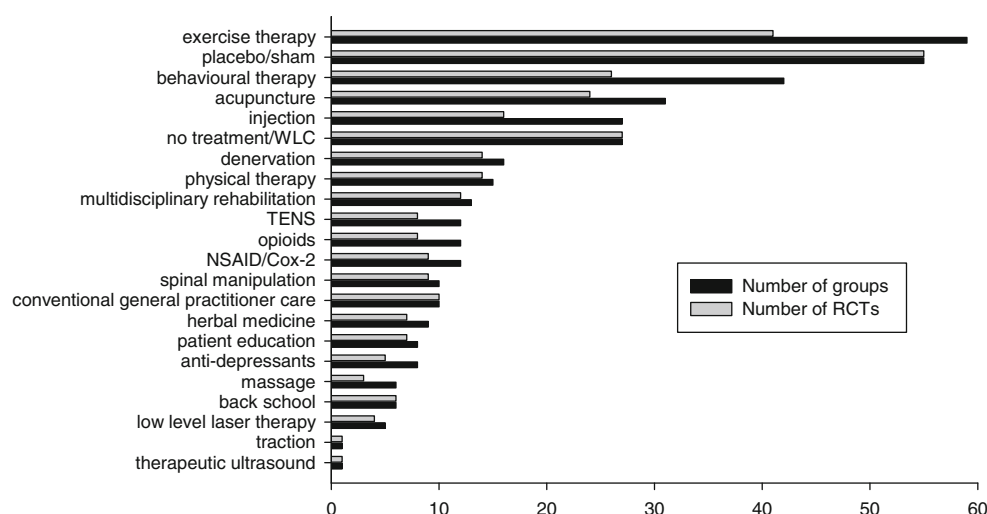
Descriptive trial characteristics, including the year of publication, sample size and interventions evaluated, were extracted from all the eligible RCTs and entered into a database for analysis. The overall risk of bias score, as well as the score per individual validity criterion, was extracted from the systematic reviews. For the purpose of analysis, the RCTs were grouped into ten time epochs with a similar number of trials in each. This allowed the identification of criteria that are poorly fulfilled and detection of changes in risk of bias (overall and per criterion) in the RCTs over time.

Results

From the four systematic reviews, the full-text versions of 157 eligible RCTs were collected and their study characteristics were extracted. The largest number of trials came from the review of physical and rehabilitation interventions ($n = 83$) [12], followed by complementary and alternative therapies ($n = 30$) [13], injection and denervation procedures ($n = 27$) [15], and pharmacological interventions ($n = 17$) [14]. The earliest year of publication for an included trial was 1980, with an increase in the number of RCTs per year until 2008, which is the last date that literature was searched. The systematic review on complementary and alternative therapies [13] excluded five trials from the meta-analysis because of methodological “fatal flaws”. As these flaws question the veracity of the results of these trials, the risk of bias scores were not included in the current analysis.

The included RCTs were performed mostly in the USA (26%), the UK (14%), Germany (10%), the Netherlands (9%) and Australia (7%). Most of the RCTs (68%) compared two groups (i.e. intervention vs. control; intervention 1 vs. intervention 2), 24% compared three groups, and less than 8% evaluated four or five groups. Overall, there were 383 intervention or control groups evaluated in the 157 RCTs (Fig. 1). Exercise therapy ($n = 59$) and placebo/sham ($n = 55$) groups were the most common, followed by behavioural therapy ($n = 42$), acupuncture ($n = 31$), injection therapy ($n = 27$), and a no treatment or waiting list control group ($n = 27$). Some interventions, such as mechanical traction and therapeutic ultrasound, were evaluated in only one intervention group of an RCT.

The mean sample size in the RCTs was 141, with a range from 17 to 3,093 and a median of 70 participants (Table 1). There were only two trials which recruited more than 1,000 participants, both of which evaluated

Fig. 1 Number of intervention and control groups evaluated in 157 RCTs from 1980 to 2008**Table 1** Number of RCTs, their sample sizes and risk of bias items in each time epoch

Year of publication	Number of RCTs	Sample size		Risk of bias			High risk of bias ^a (%)
		Mean (SD)	Range	Mean (SD)	Range	Median	
1980–1989	19 (12.1%)	83 (103)	17–476	3.4 (1.8)	1–7	3	84
1990–1995	21 (13.4%)	75 (60)	20–293	4.8 (2.5)	1–11	4	67
1996–1999	16 (10.2%)	94 (50)	25–197	5.9 (2.6)	2–11	6	41
2000–2001	16 (10.2%)	113 (80)	24–254	6.2 (2.0)	3–10	6	38
2002–2003	17 (10.8%)	161 (169)	37–690	6.4 (1.8)	4–10	6	29
2004	16 (10.2%)	94 (93)	22–338	5.4 (2.4)	2–9	5.5	50
2005	11 (7.0%)	145 (186)	50–653	6.6 (2.1)	4–10	6	27
2006	13 (8.2%)	438 (821)	26–3093	4.9 (1.9)	2–9	4	67
2007	14 (8.9%)	197 (288)	20–1162	5.8 (2.4)	1–10	5.5	50
2008	14 (8.9%)	100 (99)	20–386	5.1 (1.9)	2–8	6	43
Total	157 (100.0%)	141 (278)	17–3093	5.4 (2.3)	1–11	5	51

^a “High” risk of bias defined as fulfilling <6 of the 11 items [19]

acupuncture and were conducted in Germany [16, 17]. Because of their size ($n = 3,093$; $n = 1,162$) the studies were considered outliers and when excluded from the analyses, the mean sample size decreased to 115 participants (range 17–719). The trend since 1980 of the mean sample size per RCT and per intervention or control group, minus the outlier, is shown in Fig. 2.

There was a slight increase in the mean number of risk of bias criteria fulfilled from trials published prior to 1995 to those published after 1996 (Table 1). The median (range) number of criteria met increased from 3 (1–7) in 1980–1989 to 6 (2–11) in 1996–1999. The proportion of trials with a high risk of bias (defined as fulfilling <6 of the 11 items) was inconsistent over time and overall, 51% of the included 157 RCTs were defined as having a high risk of bias. The proportion of RCTs addressing some of the individual risk of bias criteria did appear to increase over

time, notably the reporting of adequate randomisation sequence. The trend over time in the proportion of RCTs fulfilling each criterion is presented in Fig. 3. Overall, the risk of bias criteria that were poorly met were blinding of the care provider (22%), blinding of patients (42%), or blinding of outcome assessors (43%), ensuring co-interventions are comparable between groups or avoided (30%), reporting of compliance (37%), concealment of treatment allocation (38%), and performing an “intention-to-treat” analysis (42%).

When the results for the three risk of bias items about blinding (patient, care provider, outcome assessor) were stratified into medical (i.e. pharmacological or injection therapy) and non-medical interventions (i.e. physical, rehabilitation, or complementary therapy), a larger discrepancy in the proportion of RCTs fulfilling these items was observed. The trend showed that in more recent years

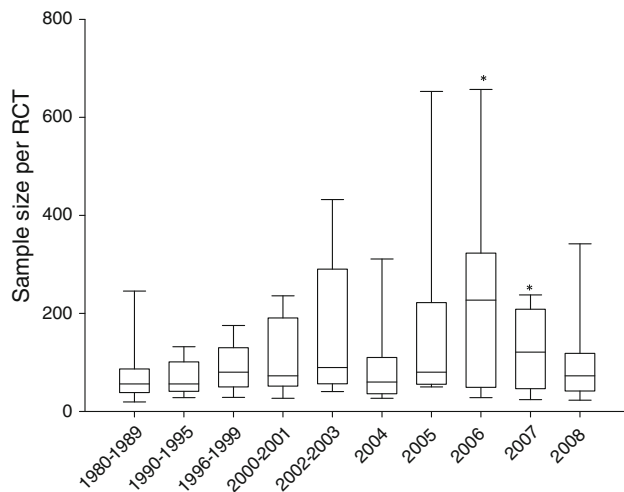


Fig. 2 Trend over time in the median sample size per RCT. Boxes represent median sample size and inter-quartile range per time category, whiskers represent 5th and 95th percentile *One outlier ($n = 3,093$) removed from 2006 [17] and one outlier ($n = 1,162$) removed from 2007 [16]

RCTs of medical interventions were more likely to be successfully blinded than RCTs of non-medical interventions (Fig. 4).

Discussion

This study evaluated the study design characteristics and risks of bias in RCTs of interventions for chronic LBP over the past 30 years. It is difficult to observe any obvious trends towards improved methodology or reporting in these trials. Despite the increasing awareness of the need for large, high-quality studies, most RCTs of interventions for chronic LBP appear to be continuing with small sample sizes evaluating complex treatments such as exercise therapy or behavioural therapy.

The 157 RCTs included in this evaluation recruited a median sample size of 70 participants, or 34 per treatment group. In practice, the statistical power afforded by this sample size would be lowered further if losses to follow-up are considered. The relatively small sample sizes observed highlights the concern that many chronic LBP trials have a high likelihood of Type II error, resulting from inadequate power to detect small but clinically relevant differences [5, 11, 18]. The sample size of an individual trial becomes less of an issue when several RCTs are pooled in a meta-analysis; however, this requires sufficient homogeneity among the interventions being evaluated as well as adequate reporting of outcomes. There was, however, a small

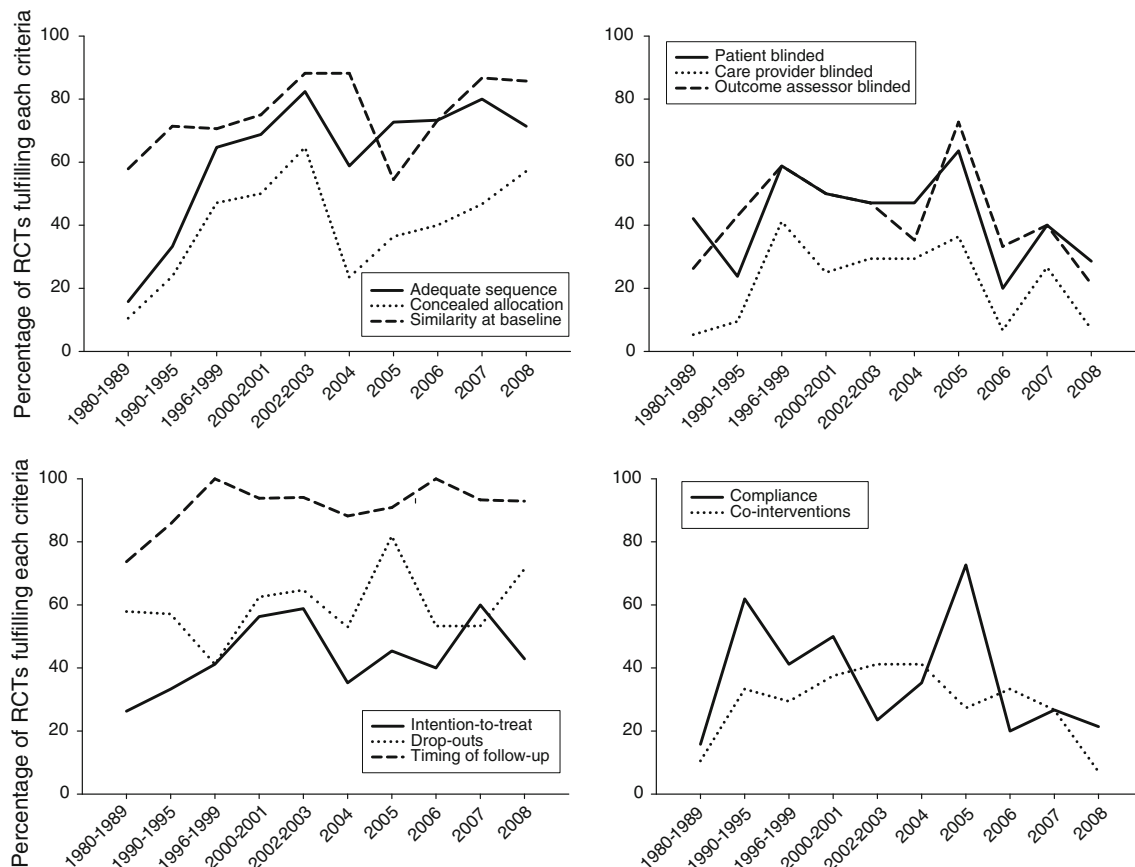


Fig. 3 Trend over time in the proportion of RCTs addressing individual items of the risk of bias criteria

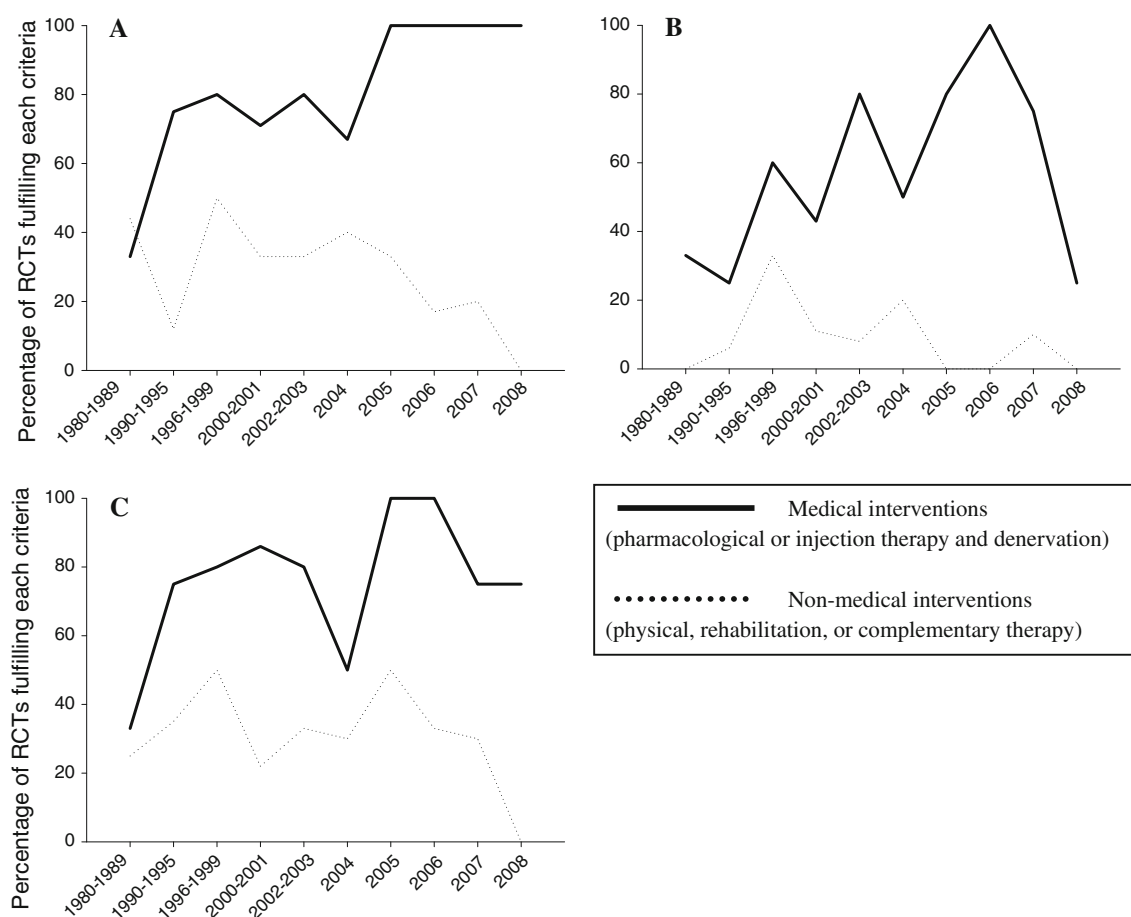


Fig. 4 Trend over time in risk of bias items on blinding **a** patient; **b** care provider; **c** outcome assessor; for RCTs on medical and non-medical interventions for chronic low-back pain

increase in the mean sample size with more recent year of publication, as well as the publication of two RCTs with over 1,000 participants in 2006 and 2007. These results are comparable with another study of 519 RCTs published in PubMed journals, where the median number of participants per treatment group, on a range of diseases and interventions, increased from 19 in 1980, to 23–39 seen from 1976 to 1991 [18].

While there was an exponential increase in the number of RCTs published per year over the last decades, the number of risk of bias criteria fulfilled appeared to increase only slightly over time. The results of this study show that the available RCTs vary greatly in risk of bias and, therefore, in their ability to provide valid and interpretable results. Apart from generating an adequate randomisation sequence, the proportion of studies fulfilling most risk of bias criteria remained steady over time. There is empirical evidence that reports of RCTs for LBP interventions consistently report smaller effect sizes if they fulfil most (i.e. >5 from 11) of the individual risk of bias criteria recommended by the CBRG. A recent study found that

trials with a higher risk of bias reported effect sizes that were, on average, 50% greater than estimates reported from trials with a lower risk of bias [19]. Further research should explore at which point the risk of bias renders the results of a study implausible, as this may justify exclusion from systematic reviews or meta-analyses. One potential limitation of the current study was that we used the risk of bias assessment which was performed by different review authors in each of the four original systematic reviews. The interpretation of risk of bias and the scoring of these items can vary across reviewers, however, all review authors were experienced in performing the risk of bias assessment as per the criteria recommended by the Cochrane Back Review Group.

In contrast to the results of this study, Maher et al. [20] reported that the methodological quality of physical therapy RCTs in the PEDro database (<http://www.pedro.org.au>) has improved over time. In this database an extensive number of studies ($n = 1,037$) dealing with lumbar spine, sacroiliac joint, or pelvic treatment is available. Although the evaluation of physical therapy RCTs included many

more trials than the current review, these differences may have arisen from inherent risks of bias within the chronic LBP trials. The difficulty in blinding patients and care providers to common non-medical interventions for chronic LBP, such as exercise, spinal manipulation, or behavioural therapy can be seen as one of these risks. The lack of a credible placebo for these interventions hampers the ability of many RCTs to fulfil the risk of bias criteria on blinding, especially for the patient and care provider. One potential way to overcome this issue would be to modify the operationalisation of the risk of bias items relating to blinding. Because of the lack of strong evidence supporting one intervention over another for chronic LBP, evaluating the treatment credibility or patient expectations and treatment preferences may provide a proxy measure for blinding in these cases. Further research on this issue is warranted to provide a clearer understanding of the methodological features which contribute to the risk of bias within a RCT for chronic LBP. In addition, exploring the reasons for differences in risk of bias between trials of different interventions (such as between medical and non-medical interventions) could contribute to explaining the large variability seen. Nevertheless, authors of chronic LBP trials should be reminded of the need to report patient compliance with the intervention being studied, the presence of co-interventions affecting participants during the study period, and the need to perform analyses using intention-to-treat principles.

One of the systematic reviews [13] excluded five studies from the meta-analysis because of “fatal flaws”. These “fatal flaws” are often considered to be characteristics of randomised controlled trials which question the veracity of the results and thus warrant ignoring the trial in decision making. Our objective was to assess trends in the risk of bias items and study methodology among trials which contribute to estimates of efficacy and decision making. Therefore, we did not include these studies in the current analysis. Further studies are needed to establish what defines a “fatal flaw” and how they can be prevented in randomised trials.

In summary, over the past 30 years, the number of RCTs for chronic LBP has increased exponentially. However, there does not appear to be a corresponding increase in sample size nor a decrease in the risk of bias. The continuing uncertainty regarding the efficacy of many interventions for chronic LBP again stresses the need for large, high quality RCTs. In addition, further research is needed into specific risks of bias within the RCTs for chronic LBP and the effect they have on the plausibility of the results.

Acknowledgments NH is supported by a fellowship from the National Health and Medical Research Council of Australia.

Conflict of interest None.

References

1. Dagenais S, Caro J, Haldeman S (2008) A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J* 8(1):8–20
2. Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, Castel LD, Kalsbeek WD, Carey TS (2009) The rising prevalence of chronic low back pain. *Arch Intern Med* 169(3):251–258
3. Koes BW, van Tulder M, Lin CW, Macedo LG, McAuley J, Maher C (2010) An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *Eur Spine J* [epub ahead of print]
4. Deyo RA, Mirza SK, Turner JA, Martin BI (2009) Overtreating chronic back pain: time to back off? *J Am Board Fam Med* 22(1):62–68
5. Koes BW, Malmivaara A, van Tulder MW (2005) Trend in methodological quality of randomised clinical trials in low back pain. *Best Pract Res Clin Rheumatol* 19(4):529–539
6. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357(9263):1191–1194
7. Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340:c332
8. Moher D, Jones A, Lepage L (2001) Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 285(15):1992–1995
9. van Tulder M, Furlan AD, Bombardier C, Bouter L (2003) Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 28(12):1290–1299
10. Furlan AD, Pennick V, Bombardier C, van Tulder M (2009) 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine* 34(18):1929–1941
11. Koes BW, Bouter LM, van der Heijden GJ (1995) Methodological quality of randomized clinical trials on treatment efficacy in low back pain. *Spine* 20(2):228–235
12. van Middelkoop M, Rubinstein SM, Kuijpers T, Verhagen AP, Ostelo R, Koes BW, van Tulder MW (2010) A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. *Eur Spine J* [epub ahead of print]
13. Rubinstein SM, van Middelkoop M, Kuijpers T, Ostelo R, Verhagen AP, de Boer MR, Koes BW, van Tulder MW (2010) A systematic review on the effectiveness of complementary and alternative medicine for chronic non-specific low-back pain. *Eur Spine J* 19(8):1213–1228
14. Kuijpers T, van Middelkoop M, Rubinstein SM, Ostelo R, Verhagen A, Koes BW, van Tulder MW (2010) A systematic review on the effectiveness of pharmacological interventions for chronic non-specific low-back pain. *Eur Spine J* [epub ahead of print]
15. Henschke N, Kuijpers T, Rubinstein SM, van Middelkoop M, Ostelo R, Verhagen A, Koes BW, van Tulder MW (2010) Injection therapy and denervation procedures for chronic low-back pain: a systematic review. *Eur Spine J* 19(9):1425–1449
16. Haake M, Muller HH, Schade-Brittinger C, Basler HD, Schafer H, Maier C, Endres HG, Trampisch HJ, Molsberger A (2007) German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. *Arch Intern Med* 167(17):1892–1898

17. Witt CM, Jena S, Selim D, Brinkhaus B, Reinhold T, Wruck K, Liecker B, Linde K, Wegscheider K, Willich SN (2006) Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *Am J Epidemiol* 164(5):487–496
18. Chan AW, Altman DG (2005) Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 365(9465):1159–1162
19. van Tulder MW, Suttrop M, Morton S, Bouter LM, Shekelle P (2009) Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine* 34(16):1685–1692
20. Maher CG, Moseley AM, Sherrington C, Elkins MR, Herbert RD (2008) A description of the trials, reviews, and practice guidelines indexed in the PEDro database. *Phys Ther* 88(9):1068–1077