Review

A Systematic Review and Meta-Analysis on the Effectiveness of Graded Activity and Graded Exposure for Chronic Nonspecific Low Back Pain

Ibai López-de-Uralde-Villanueva, PT, MSc,[*†‡§]
Daniel Muñoz-García, PT, MSc,[*†]
Alfonso Gil-Martínez, PT, MSc,[*†‡§]
Joaquín Pardo-Montero, PhD,[*†‡§]
Rosa Muñoz-Plata, PT, MSc,[*†]
Santiago Angulo-Díaz-Parreño, MSc,[†‡¶]
Miguel Gómez-Martínez, OT, PhD,[†**]
and Roy La Touche, PT, MSc, PhD[*†‡§]

*Department of Physiotherapy, Faculty of Health Science, The Center for Advanced Studies University La Salle, Universidad Autónoma De Madrid, Aravaca, Madrid, Spain; [†]Motion in Brains Research Group, The Center for Advanced Studies University La Salle, Universidad Autónoma De Madrid, Aravaca, Madrid, Spain; [‡]Institute of Neuroscience and Craniofacial Pain (INDCRAN), Madrid, Spain; [§]Hospital La Paz Institute for Health Research, IdiPAZ, Madrid, Spain; [¶]Faculty of Medicine, Universidad San Pablo CEU, Madrid, Spain; [**]Department of Occupational Therapy, Faculty of Health Science, The Center for Advanced Studies University La Salle, Universidad Autónoma De Madrid, Aravaca, Madrid, Spain

Reprint requests to: Roy La Touche, PT, MSc, PhD, Facultad de Ciencias de la Salud Centro Superior de Estudios Universitarios La Salle, Calle la Salle, 10, 28023 Madrid, Spain. Tel: + 34 91 7401980 (EXT. 256); Fax: +34-91-3571730; E-mail: roylatouche@yahoo.es.

Abstract

Objective. Our aim was to systematically review and meta-analyze the effectiveness of graded activity (GA) or graded exposure (GEXP) for chronic nonspecific low back pain (CNSLBP).

Methods. A literature search of multiple databases (MEDLINE, EMBASE, PEDro, CINAHL, and PsychINFO) was conducted to identify randomized control trials (RCTs). Standardized mean difference (SMD) and 95% confidence intervals (CI) were calculated for relevant outcome measures (pain intensity, disability, quality of life, and catastrophizing).

Results. Thirteen RCTs met the inclusion criteria. Only nine studies were included in the meta-analysis. GA was significantly more effective than the control group (CG) for improvements in disability in the short term (three studies: $n = 254$, SMD $= -0.3$, 95% CI -0.55 to -0.05, $P = 0.02$) and long term (two studies: $n = 238$, SMD $= -0.53$, 95% CI -0.79 to -0.27, $P < 0.0001$). GA was significantly less effective than GEXP for the improvement of disability in the short term (two studies: $n = 105$, SMD $= 0.39$, 95% CI 0.003–0.78, $P = 0.048$). GA was also significantly less effective than GEXP at improving catastrophizing in the short term (two studies: $n = 105$, SMD $= 0.48$, 95% CI 0.09–0.87, $P = 0.02$).

Conclusion. Limited evidence has been found to show that GA significantly reduces disability in the short and long term when compared with the CG in CNSLBP. There is moderate evidence that GEXP more effectively decreases catastrophizing than GA in the short term. No difference was found between GA and other exercise for any variable.

Key Words. Recurrent Low Back Pain; Pain Catastrophizing; Cognitive Behavior Therapy; Behavioural Graded Activity; Physiotherapy

Introduction

Low back pain is one of the major health problems of Western society resulting in serious economic and social

costs [1]. The annual prevalence of low back pain in the general population is 38%, with a higher incidence among middle age women and with a three-month minimum episode duration (chronic) in 20% of the cases [2].

Some results of chronic nonspecific low back pain (CNSLBP) are mobility restriction, long-term disability, and decreased quality of life; additionally, this is one of the main causes of sick leave [3]. The literature suggests that psychosocial factors, such as catastrophizing, kinesiophobia, or depression could influence the development and perpetuation of pain and also decrease the level of activity as the fear avoidance model predicts [4,5]. Recent scientific evidence shows that there is an association between psychosocial factors and disability [6,7]. Psychosocial factors may be associated with poor recovery in patients with low back pain [8]. In addition, improving psychosocial factors involved in CNSLBP has been shown to reduce the perception of disability [9].

Therapeutic exercise is often used in the treatment of CNSLBP [10]. The efficacy of exercise on pain and disability levels in patients with CNSLBP has been clearly demonstrated [11,12], but the most appropriate dose and type is still unknown [13].

There are challenging interventions based on exercise for CNSLBP using a cognitive behavioral approach to improve activity tolerance, labeled as graded activity (GA) and graded exposure (GEXP) [14]. Both interventions have several features in common: they aim to restore functionality by decreasing patient disability and increasing patient education as well as setting feasible goals of specific behaviors. The main differences between treatments are that GA aims to achieve this target by positively reinforcing patient activity levels [15], while the GEXP approach encourages a confrontation response by exposing patients to specific situations of which they are fearful during rehabilitation [16]. GA consists of three phases: measuring functional capacity, educating in the workplace and providing an individual program of submaximal exercise that is gradually increased [17].

Previous systematic reviews have suggested that GA is no more effective than other forms of exercise in patients with persistent low back pain [14,18]. However, these reviews included patients with acute and subacute pain, whereas the current review focused on chronic, nonspecific conditions. In any case, GA does serve as a good contrast for GEXP [16]. In addition, systematic reviews published to date only include studies published before 2009 [14] and 2011 [18].

Therefore, the aim of this study was to systematically review and meta-analyze the randomized clinical trials (RCTs) evaluating the effectiveness of GA and GEXP over other interventions for pain, disability, and psychosocial factors associated with CNSLBP patients.

Methods

The systematic review and meta-analysis was performed in accordance with the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines [19]. The PRISMA statement is composed of a 27-item checklist and a four-phase flow diagram, which assists in reporting systematic reviews and meta-analyses [19].

Inclusion Criteria of the Studies

The selection criteria used in this review are based on methodological and clinical aspects such as the population, intervention, control, outcomes, and study design [20] criteria defined for this systematic review as follows:

Population

The patients selected for the trials had to be over 18 years old and had CNSLBP for at least 3 months. CNSLBP is defined as pain that is localized between the costal margins and the inferior gluteal folds, with or without leg pain, that does not have a specified physical cause (e.g., trauma, neoplasm, nerve root compression, osteoporosis, infection, or inflammatory rheumatic processes) [1,21].

Intervention and Control

The studies that were included had to compare GEXP, behavioral GA, or GA to usual care (e.g., treatment of the medical specialist or usual physiotherapy; this last consists of joint manipulation/mobilization, massage, back-care advice, and/or recommendation to active approach not supervised), and receive no treatment or other interventions based on therapeutic exercise or when GA or GEXP was added as a supplement to other interventions.

The studies were considered to have evaluated GA when the treatment included the following features: a) treatment goals were functional activities; and b) the therapeutic exercise program was based on operant-conditioning behavioral principles and improving exercise and activity tolerance using a quota system instead of pain abatement.

The studies were considered to have evaluated GEXP when the treatment included the following features: a) feared exercise or activities were identified; and b) exposure proceeded in a hierarchical fashion, starting with an exercise or activity that elicited minimal amounts of fear and progressing only when anxiety lessened.

Outcomes

The measures used to check the results and effects of treatment had to assess at least two or more of the related variables of pain intensity, disability, pain catastrophizing, and quality of life measures. The pain intensity had to be assessed with a visual analog scale (minimal clinically important change, MCIC = 15 mm) [22], a Numerical Rating Scale [23] (MCIC = 2-point) [24], or the McGill Pain Questionnaire (standard- or short-version) [25,26]; these McGill Pain Questionnaires used a visual analog scale to calculate the pain intensity score. Disability was considered if was assessed with the 24-Item Roland-Morris Disability Questionnaire [27] (MCIC = 5-point) [22], the Quebec Back Pain Disability Scale [28] (MCIC = 20-point) [22], or the Pain Disability Index [29] (MCIC = no published). The pain catastrophizing had to be evaluated with the Pain Catastrophizing Scale [30] (minimum detectable change, MDC = 9.1) [31], or the Pain Cognition List-pain catastrophizing [32] (MCIC = no published), whereas the quality of life was considered when was assessed with the EuroQoL Questionnaire [33] (MCIC = 0.03) [34], or the 36-Item Short-Form Health Survey Questionnaire [35] (MCIC for the physical component summary = 4.9-point) [36]. All these variables have shown an adequate validity and reliability [23,25–28,32,35,37–45]. Furthermore, these had to be registered in the short term (<3 months), intermediate term (between 3 months and 12 months) or long term (>12 months) follow-up proposed by the Cochrane Back Review Group [46].

Study Design

RCTs were selected. No restrictions were applied to any specific language as recommended by the international criteria [47].

Search Strategy

The search of scientific articles was performed using MEDLINE (1950 to December 2013), EMBASE (1988 to December 2013), PEDro (1999 to December 2013) CINAHL (1982 to December 2013), and PsychINFO (1806 to December 2013), with an end date of December 10, 2013.

The terms used for the search were derived from a combination of the following medical subject headings (MeSH) and other non-MeSH terms. The terms and Boolean operator used in MeSH were: "low back pain OR back pain OR backache," "exercise therapy," "behavior therapy," "cognitive therapy," and "physical therapy modalities." The Boolean operator "AND" connected the intervention and the patient group. As additional filters, randomized controlled trial (in "article types") humans (in "species") and adult (in "ages") were chosen. The non-MeSH terms used were: "nonspecific chronic low back pain," "chronic low back pain,"

"graded activity," "graded exposure," "behavioral graded activity," "gradual exercises," and "behavioral therapy." The search strategy was adapted for each database as necessary. Two independent reviewers conducted the search using the same methodology, and the differences that emerged in this phase were resolved by consensus.

Selection Criteria and Data Extraction

First, an analysis of data was performed by two independent reviewers (I.L.U.V. and D.M.G.) who assessed the relevance of RCTs regarding the studies' question and objectives. This first analysis was made based on information from the title, abstract and keywords of each study. If there were no consensus or the abstracts did not contain sufficient information, it was agreed to review the full text.

In the second phase of the analysis using the full text, we proceeded to test whether the studies met all of the inclusion criteria. Differences between reviewers were resolved by a process of discussion/consensus moderated by a third reviewer (R.L.T.) [46]. Data described in the results were extracted by means of a structured protocol that ensured that the most relevant information was obtained from each study [48].

Methodological Quality Assessment

The assessment of the methodological quality of the studies was performed using the PEDro scale [49], which is based on the Delphi list [50]. The reliability of this scale has been demonstrated by Tooth et al. [51] and Maher et al. [49].

The PEDro scale assesses the internal and external validity of a study and consists of 11 criteria: 1) specified study eligibility criteria, 2) random allocation of subjects, 3) concealed allocation, 4) measure of similarity between groups at baseline, 5) subject blinding, 6) therapist blinding, 7) assessor blinding, 8) fewer than 15% dropouts, 9) intention-to-treat analysis, 10) between-group statistical comparisons, and 11) point measures and variability data. Criteria (2)–(11) were used to calculate the PEDro score. The methodological criteria were scored as: *Yes* (one point), *No* (zero points) or *Don't know* (zero points). The PEDro score of each selected study provided an indicator of the methodological quality (9–10 = excellent; 6–8 = good; 4–5 = fair; <4 = poor) [52].

Two independent reviewers (I.L.U.V. and D.M.G.) examined the quality of all of the selected studies using the same methodology; disagreements between reviewers were resolved by consensus including a third reviewer (R.L.T.). All reviewers had at least 5 years of experience in health sciences research and were trained to critically appraise studies using the PEDro scale criteria. The inter-rater reliability was determined using Kappa coefficient (>0.7 means high level of agreement between

assessors, between 0.5 and 0.7 a moderate level of agreement, and <0.5 a low level of agreement) [53]. Consistency between the two reviewers who performed the methodological assessment was evaluated with the Cronbach's coefficient α.

Qualitative Analysis

The qualitative analysis used is based on the classification of the results in levels of evidence according to Van Peppen et al.'s model [54]. Evidence was categorized into the following five levels depending on the methodological quality.

1. Strong evidence—provided by statistically significant findings in outcome measures in at least two high-quality RCTs, with PEDro scores of at least four points.
2. Moderate evidence—provided by statistically significant findings in outcome measures in at least one high-quality RCT and at least one low-quality RCT (≤3 points on PEDro) and/or one high-quality clinical controlled trial (CCT).
3. Limited evidence—provided by statistically significant findings in outcome measures in at least one high-quality RCT and/or at least two high-quality CCTs (in the absence of high-quality RCTs).
4. Indicative findings—provided by statistically significant findings in outcome measures in at least one high-quality CCT and/or low-quality RCTs (in the absence of high-quality RCTs), and/or two studies of a nonexperimental nature with sufficient quality (in the absence of RCTs and CCTs).
5. No or insufficient evidence—in cases where the results of eligible studies do not meet the criteria for one of the above-stated levels of evidence, and/or in the case of conflicting (statistically significant positive and statistically significant negative) results among RCTs and CCTs, or in the case of a lack of eligible studies.

If the number of studies that showed evidence is <50% of the total number of studies found within the same category of methodological quality and study design (RCTs, CCTs, or nonexperimental studies), no evidence was classified [54].

Data Synthesis and Analysis

Statistical analysis was conducted using meta-analysis with interactive explanations (MIX, version 1.7) [55]. The meta-analysis was performed in accordance with the PRISMA guidelines [19].

To provide a comparison between outcomes reported by the studies, the standardized mean difference (SMD) over time and corresponding 95% confidence interval (CI) were calculated for continuous variables, if possible, and short-term, intermediate-term, and long-term follow-up time points. The statistical significance of the pooled SMD was examined by the Z-test. The same

inclusion criteria were used for the systematic review, as well as for the meta-analysis, and four more criteria were included: 1) In the results, there was detailed information regarding the comparative statistical data of the exposure factors, therapeutic interventions, and treatment responses; 2) The GA or GEXP treatment was compared with a control group (CG) or another treatment; 3) Data of the analyzed variables were represented in at least two studies; and 4) A minimum score of six points (PEDro) was considered necessary for inclusion in the meta-analysis, as inclusion of lower quality studies in a meta-analysis may overestimate the treatment effect of interventions [47].

The estimated SMDs were interpreted as described by Hopkins et al. [56]; that is, a SMD of 4.0 was considered to represent an extremely large clinical effect, 2.0–4.0 a very large effect, 1.2–2.0 a large effect, 0.6–1.2 a moderate effect, 0.2–0.6 a small effect, and 0.0–0.2 a trivial effect.

To compare and contrast data, the following statistical tests were performed: the DerSimonian-Laird Q-test to measure the level of heterogeneity and Egger's regression tests to examine publication bias [57]. When the Q-test was significant ($P < 0.05$), this indicated that heterogeneity existed among the studies. Therefore, a random-effects model was used in the meta-analysis.

Results

The study search strategy is shown in the form of a flow chart (Figure 1). Finally, we selected 12 articles that met the inclusion criteria. Table 1 presents the general descriptive characteristics of these studies.

Characteristics of the Study Population

All patients in the studies had nonspecific low back pain for at least 3 months. The required research designs were randomized controlled studies and the samples were varied; the total number of patients included was 1,486. The percentage of women participating in the studies ranged from 42% [60] to 75% [66], excepting two studies in which this information was not specified [59,68].

The average age of participants in the included studies ranged from 35 [66] to 49 [61] years of age. The percentage of withdrawals and dropouts during treatment in the studies ranged from 0% [59,66] to 47% [68].

Study Characteristics

The characteristics for which data were extracted (sample size, participants ages, duration of complaint, study design, intervention, outcomes, follow-up period, and main results) are presented in Table 1.

Six studies included in this systematic review compared GA to other forms of exercise [58,61–65]. The interventions involved in other forms of exercise were motor control (addressed to improve activity of muscles with
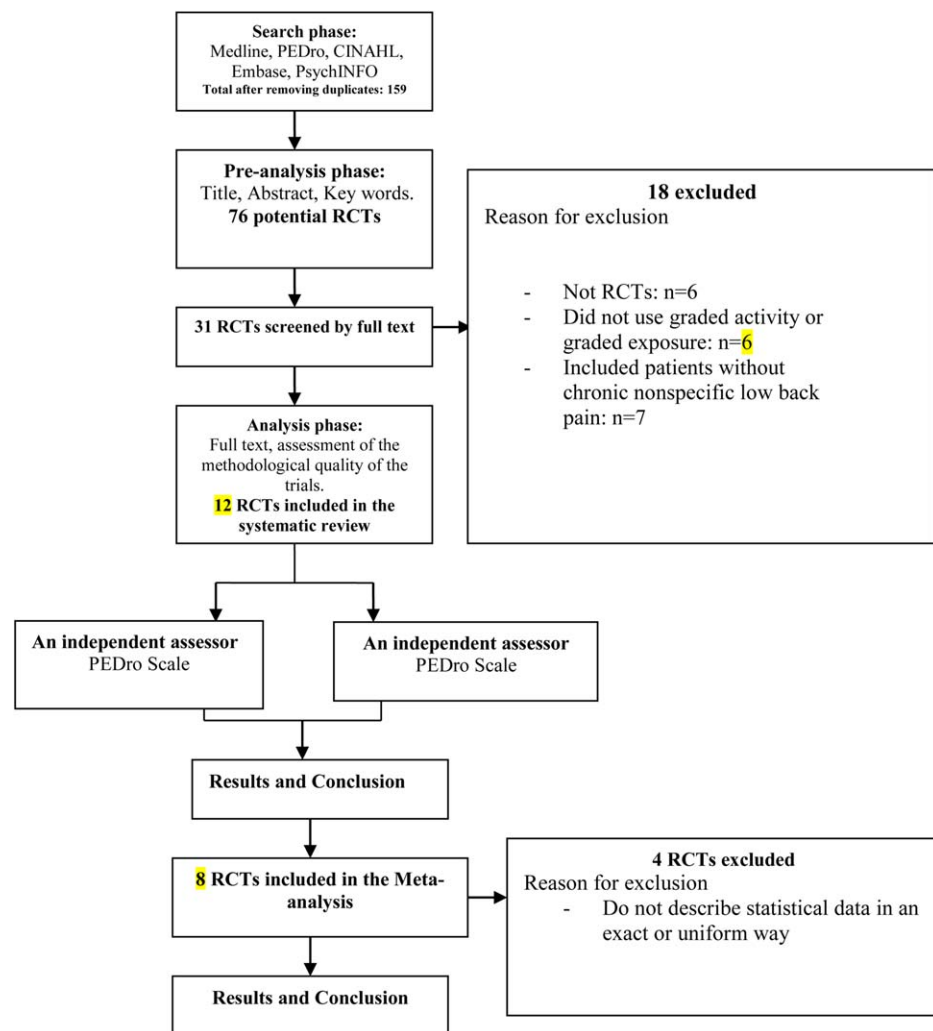
Figure 1 Flowchart of the selection of the clinical trials. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

poor control and reducing activity of any overactive muscle, thereby enhancing the spinal stability) [58,61], active physical (aerobic training on a bicycle, training of strength and endurance for low back and upper leg muscles) [62–64], and physical therapy according to clinical guidelines (advising to stay active and providing exercise therapy supervised by a physiotherapist) [65].

In six studies, GA was compared with a CG [58,60,62–64,68]. The following interventions were considered in the CG: wait list [62–64,68], usual care (treatment applied by medical specialist, occupational physician, general practitioner, and/or allied health professionals) [58,60].

No adverse effects were reported in any of the studies included in this systematic review.

Methodological Quality Analysis

The RCTs were evaluated with the PEDro scale and a median score of 5.9 2 ± 1.78 (Mean ± SD; range: 3–8)

was revealed. According to analyses of the two reviews, eight of the studies had methodologies that were good in terms of quality. Table 2 shows the results of the evaluation according to the PEDro scale. The two reviewers reported a discrepancy in the evaluation of four RCTs; the discrepancy for those studies concerned their scores for Items 3, 7, and 11 (3: concealed allocation; 7: assessor blinding; 11: point measures and variability data). A consensus was reached after a third reviewer intervened. The inter-rater reliability of the methodological quality assessment was high ($\kappa = 0.82$, 95% CI 0.65–0.93). Most of the RCTs lacked blinding in the assessor, care provider, and patients. When the PEDro scores of the two reviewers were compared, consistency was high ($\alpha = 0.84$).

Systematic Review Results

Graded Activity vs Other Forms of Exercise

Six studies (1,167 patients) compared GA to other forms of exercise [58,61–65]. No statistically significant

Table 1 Characteristics of the included studies (N = 12)

Trial	N	Participant Ages, and Duration of Complaint	Study Design and Interventions	Outcomes (Measures)	Follow-Up	Results/Authorś Conclusion
Critchley et al. 2007 [58]	212	Age: >18 years Duration: >3 months	RCT EG1: Motor control (exercise) EG2: GA CG: usual outpatient physiotherapy	Pain (NRS) Disability (RMDQ-24) Quality of life (EQ-5D)	6 months 12 months 18 months	Three types of physiotherapy produced clinical improvements in health status. The pain coping focus physical therapy decrease medical service frequentation rates and shows better cost-effectiveness.
De Jong et al. 2005 [59]	6	Age: 18–65 years Duration: ≥6 months	RCT EG1: GEXP EG2: GA	Pain (VAS) Disability (RMDQ-24)	Post-treatment (6 or 8 weeks) 6 months	Patient therapeutic education significative decreases en kinesiophobia, catastrophizing, and fear of pain. GEXP: greater effects and maintained over time. GA only maintains therapeutic education improvements.
Lambeek et al. 2010 [60]	134	Age: 18–65 years Duration: >3 months	RCT EG1: GA CG: usual care (medical specialist)	Pain (VAS) Disability (RMDQ-24)	3 months 6 months 12 months	EG2 reduces job and personal disability, Nevertheless there were no effects in pain reducing.
Leeuw et al. 2008 [15]	85	Age: 18–65 years Duration: >12 weeks	RCT EG1: GEXP EG2: GA	Pain [MPQ (VAS)] Disability (QBPDS) Catastrophizing (PCS)	Post-treatment (8 or 13 weeks) 6 months	6 months later GEXP was effective in chronic LBP, but not mote tan GA. GEXP more significative changes in catastrophism and perceiving painful activities.
Macedo et al. 2012 [61]	172	Age: 18–80 years Duration: >3 months	RCT EG1: motor control (exercise) EG2: GA	Pain (NRS) Disability (RMDQ-24) Quality of life (SF-36)	Post-treatment (2 months) 6 months 12 months	Similar effects of both treatments: pain and disability reduction, increasing function, global change ratings, and quality of life; patients with LBP.
Smeet et al. 2006 [62] Smeet et al. 2008 [63] Smeet et al. 2009 [64]	223	Age: 18–65 years Duration: >3 months	RCT EG1: Active physical (exercise) EG2: GA EG3: GA + Active physical (exercise) CG: wait-list	Pain (VAS) Disability (RMDQ-24) Catastrophizing (PCL-pain catastrophizing) Quality of life (EQ-5D)	10 weeks 6 months 26 weeks 12 months 52 weeks	2006 → All treatments improved catastrophizing, perceived disability, complaints, and pain intensity; CG with no effects. But not only CBT groups changes catastrophizing. 2008 → 1 year later, individual treatments were at least equal than combined treatments.

Table 1 Continued

Trial	N	Participant Ages, and Duration of Complaint	Study Design and Interventions	Outcomes (Measures)	Follow-Up	Results/Author's Conclusion
						2009 → EG2 were more cost-effective in reducing disability and improving quality of life. In lesser extent, EG1 was better than EG3 reducing disability. Individual treatments were more cost-effective than combined treatments from 10 weeks until 1 year follow-up.
Van der Roer et al. 2008 [65]	114	Age: 18–65 years Duration: >3 months	RCT EG1: GA EG2: exercise therapy supervised by a physiotherapist	Pain (NRS) Disability (RMDQ-24)	6 weeks 13 weeks 26 weeks 52 weeks	EG1: more effective reducing pain, increasing coping and self-efficacy and greater number of patients perceived improvements than CG. Nevertheless, 1 year follow-up, no statistically significant. EG1 were no more effective than controls with chronic pain patients.
Vlaeyen et al. 2001 [66]	4	Age: 18–65 years Duration: >5 years	Crossover study EG1: 1° GEXP, 2° GA EG2: 1°GA, 2° GEXP	Disability (RMDQ-24) Catastrophizing (PCL-pain catastrophizing)	2 months	Only improvements in knowledge about pain and fear (catastrophizing, pain disability, pain control) in GEXP, but not in GA group, considering treatment order.
Vlaeyen et al. 2002 [67]	7	Age: 18–60 years Duration: >24 weeks	Crossover study EG1: 1° GEXP, 2° GA EG2: 1°GA, 2° GEXP	Pain (VAS) Disability (RMDQ-24)	12 weeks 12 months	Only GEXP exhibited improvements in pain related fear or catastrophizing pain; not in GA group, regardless treatment order.
Woods et al. 2008 [68]	83	Age: 18–65 years Duration: nonspecified	RCT EG1: GEXP EG2: GA CG: wait-list	Pain [SF-MPQ (VAS)] Disability (PDI) Catastrophizing (PCS)	Post-treatment (4 weeks) 4 weeks	EG1 fear—avoidance more improvements than EG2 or Controls. There improvements continue in a month later follow-up.

EG = experimental group; CG = control group; CBT = cognitive-behavioral therapy; GEXP = graded exposure; GA = graded activity; LBP = low back pain; CNLBP = chronic non-specific low back pain; VAS = visual analog scale; NRS = numerical rating scale; MPQ = Mcgill pain questionnaire; SF-MPQ = short form-McGill pain questionnaire; RMDQ-24 = 24-Item Roland-Morris disability questionnaire; QBPDS = Quebec back pain disability scale; PDI = pain disability index; PCS = pain catastrophizing scale; PCL-pain catastrophizing = pain cognition list-pain catastrophizing; EQ-5D = EuroQoL questionnaire; SF-3 6 = 36-item short-form health survey questionnaire; RCT = randomized controlled trial.

Table 2 Methodologicl quality of the studies included in the systematic review (PEDro scores)

TRIAL	Specified Study Eligibility Criteria Item 1	Random Allocation of Subjects Item 2	Concealed Allocation Item 3	Similarity Between Groups at Baseline Item 4	Subject Blinding Item 5	Therapist Blinding Item 6	Assessor Blinding Item 7	Fewer than 15% Dropouts Item 8	Intention-to-Treat Analysis Item 9	Between-Group Statistical Comparisons Item 10	Point Measures and Variability Data Item 11	TOTAL
Critchley et al., 2007 [58]	—	1	1	1	0	0	1	0	1	1	1	7
De Jong et al., 2005 [59]	—	1	0	0	0	0	0	1	0	0	1	3
Lambeek et al., 2010 [60]	—	1	1	1	0	0	0	1	1	1	1	7
Leeuw et al., 2008 [15]	—	1	1	1	0	0	0	0	1	1	1	6
Macedo et al., 2012 [61]	—	1	1	1	0	0	1	1	1	1	1	8
Smeets et al., 2006 [62]	—	1	0	1	0	0	1	1	1	1	1	7
Smeets et al., 2008 [63]	—	1	1	1	0	0	1	1	1	1	1	8
Smeets et al., 2009 [64]	—	1	1	1	0	0	0	1	0	1	1	6
Van Der Roer et al., 2008 [65]	—	1	1	1	0	0	0	1	1	1	1	7
Vlaeyen et al., 2001 [66]	—	1	1	0	0	0	0	1	0	0	0	3
Vlaeyen et al., 2002 [67]	—	1	1	0	0	0	0	1	0	1	0	4
Woods et al., 2008 [68]	—	1	0	1	0	0	0	0	1	1	1	5

differences were found for pain intensity, catastrophizing, quality of life, and disability [58,61–65]. No study found statistically significant differences in disability and pain intensity outcomes. In addition, some studies concluded that GA showed lesser total costs (most cost-effective), although not statistically significant, than other forms of exercise or even than the combination of GA with physical activity [58,64].

Graded Activity vs Control Group

Six studies with a total of 1,098 patients compared GA to CG [58,60,62–64,68]. In most of them, there were no statistically significant differences for pain, disability, catastrophizing, and quality of life [58,60,68]. However, Smeet et al. [62–64] reported statistically significant differences for these outcomes. In the same way, Lambeek et al. [60] showed statistically significant differences in long-term disability. In addition, it is important to mention that there is a clear trend toward the greater effectiveness of GA, although this difference is not significant.

Graded Activity vs Graded Exposure

Five trials (185 patients) compared GA with GEXP and the CG [15,59,66–68]. Two of them had a crossover design; therefore, it is more difficult to draw conclusions because no study evaluated the immediate effects of every intervention [66,67]. No statistically significant differences were found for disability and pain [15,68], although there were statistically significant differences reported for catastrophizing [15]. In spite of the lack of differences, the studies showed slightly higher results for GEXP than GA.

Meta-Analysis Results

Meta-Analysis Comparison of GA vs Other Forms of Exercise

Six RCTs evaluated the effects of GA on disability compared with other forms of exercise. Meta-analysis for these studies showed no statistically significant differences in the reduction of disability in the short term (three studies: [61–63] $n = 372$, SMD $= -0.09$, 95% CI -0.3 to 0.11, $Z = 0.9$, $P = 0.37$, heterogeneity: Q-value $= 2.2$, $P = 0.33$; Figure 2A) or in the intermediate term (three studies: [58,61,63] $n = 367$, SMD $= -0.09$, 95% CI -0.29 to 0.12, $Z = 0.83$, $P = 0.41$, heterogeneity: Q-value $= 2.2$, $P = 0.33$; Figure 2B) nor the long term (four studies: [58,61,63,64] $n = 461$, SMD $= -0.12$, 95% CI -0.3 to 0.07, $Z = 1.24$, $P = 0.22$, heterogeneity: Q-value $= 3.21$, $P = 0.36$; Figure 2C). There was no evidence of publication bias for any of the three meta-analyses (short term, $P = 0.08$; intermediate term, $P = 0.18$; long term, $P = 0.07$).

GA was not found to be significantly more effective than other forms of exercise with regard to improvements in the quality of life in the long term (two studies: [58,61] $n = 254$, SMD $= 0.21$, 95% CI -0.03 to 0.46, $Z = 1.7$, $P = 0.09$, heterogeneity: Q-value $= 0.1$, $P = 0.31$; Figure 3).

Four RCTs which included the variable pain intensity were analyzed in the meta-analysis, showing no statistically significant differences in the short term (three studies: [61–63] $n = 372$, SMD $= -0.18$, 95% CI -0.38 to 0.02, $Z = 1.73$, $P = 0.08$, heterogeneity: Q-value $= 2.23$, $P = 0.33$; Figure 4A) or the intermediate term (three studies: [58,61,63] $n = 367$, SMD $= -0.02$, 95% CI -0.22 to 0.19, $Z = 0.15$, $P = 0.88$, heterogeneity: Q-value $= 1$, $P = 0.6$; Figure 4B), as well as in the long term (three studies:[58,61,63] $n = 357$, SMD $= -0.1$, 95% CI -0.31 to 0.11, $Z = 0.92$, $P = 0.36$, heterogeneity: Q-value $= 0.68$, $P = 0.71$; Figure 4C). There was no evidence of publication bias for any of the three meta-analyses (short term, $P = 0.08$; intermediate term, $P = 0.92$; long term, $P = 0.24$).

Meta-Analysis Comparison of GA vs Control Group

GA was significantly more effective than the CG with regard to improvements in disability in the short term (three studies:[60,62,68] $n = 254$, SMD $= -0.3$, 95% CI -0.55 to -0.05, $Z = 2.33$, $P = 0.02$, heterogeneity: Q-value $= 5.7$, $P = 0.06$; Figure 5A) and long term (two studies:[58,60] $n = 238$, SMD $= -0.53$, 95% CI -0.79 to -0.27, $Z = 3.97$, $P < 0.0001$, heterogeneity: Q-value $= 1.6$, $P = 0.21$; Figure 5C). There were no statistically significant differences in the intermediate term (two studies:[58,60] $n = 231$, SMD $= -0.16$, 95% CI -0.42 to 0.1, $Z = 1.23$, $P = 0.22$, heterogeneity: Q-value $= 1$, $P = 0.31$; Figure 5B), and there was no evidence of publication bias for meta-analysis in the short term ($P = 0.9$).

A total of four RCTs evaluated the effects of GA on pain intensity compared with the CG. Meta-analysis of these studies showed no statistically significant differences in the reduction of pain intensity in the short term (three studies:[60,62,68] $n = 256$, SMD $= -0.25$, 95% CI -0.77 to 0.28, $Z = 0.92$, $P = 0.36$, heterogeneity: Q-value $= 7.6$, $P = 0.02$; Figure 6A), as well as in the intermediate term (two studies:[58,60] $n = 236$, SMD $= 0.1$, 95% CI -0.16 to 0.35, $Z = 0.75$, $P = 0.46$, heterogeneity: Q-value $= 0.5$, $P = 0.48$; Figure 6B), or in the long term (two studies: [58,60] $n = 222$, SMD $= -0.17$, 95% CI -0.43 to 0.1, $Z = 1.23$, $P = 0.22$, heterogeneity: Q-value $= 0.02$, $P = 0.1$; Figure 6C). There was no evidence of publication bias for meta-analysis in the short term ($P = 0.8$).

Meta-Analysis Comparison of GA vs GEXP

GA was significantly less effective than GEXP with regard to improvements in disability in the short term (two studies: [15,68] $n = 105$, SMD $= 0.39$, 95% CI
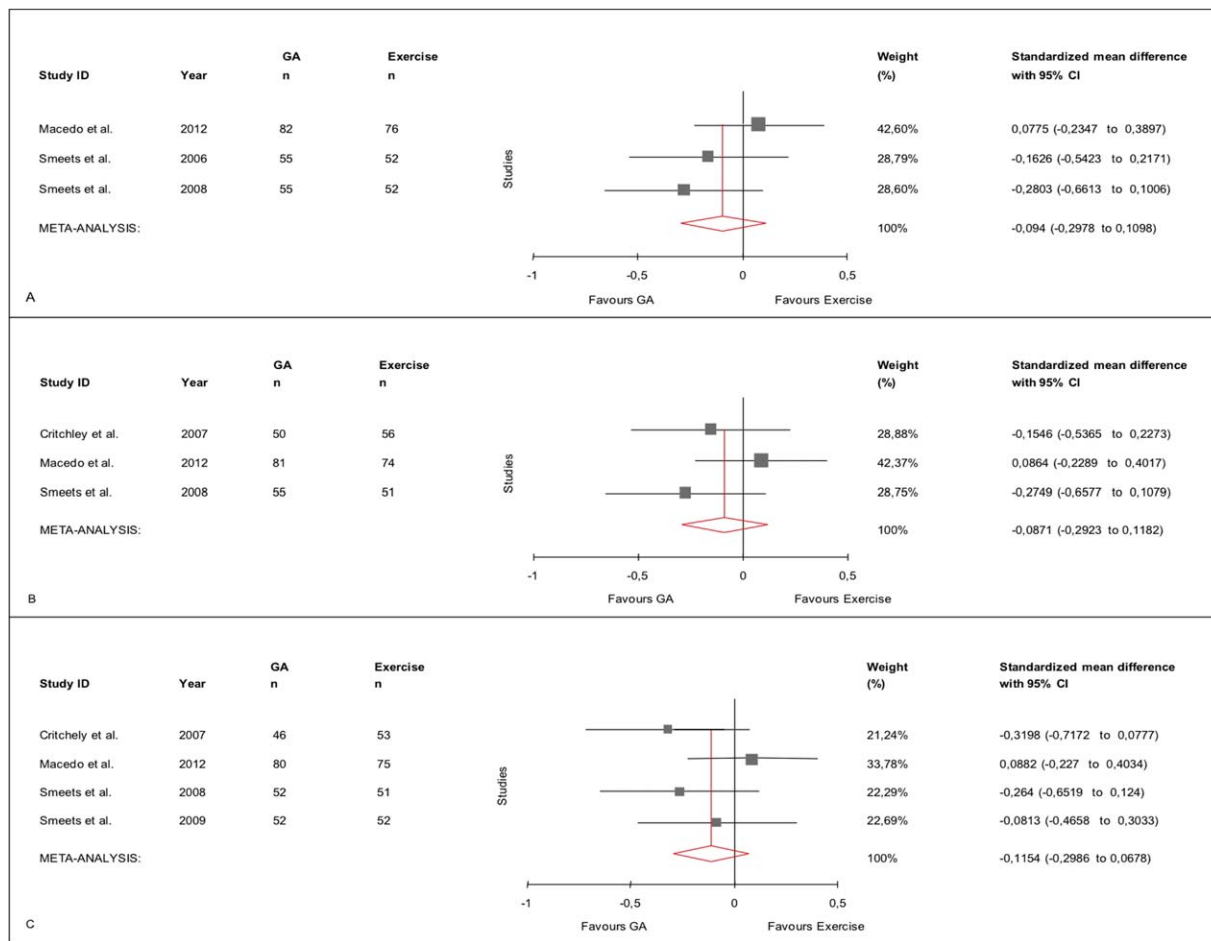
Figure 2 Forest plots of GA vs other forms of exercise effect on disability. (A): short term; (B): intermediate term; (C): long term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

0.003–0.78, $Z = 1.98$, $P = 0.048$, heterogeneity: Q-value = 1.1, $P = 0.29$; Figure 7).

GA was also significantly less effective than GEXP at improving catastrophizing in the short term (two studies: [15,68] $n = 105$, SMD = 0.48, 95% CI 0.09–0.87, $Z = 2.39$, $P = 0.02$, heterogeneity: Q-value = 1.4, $P = 0.24$; Figure 8).

Meta-analysis showed no statistically significant differences in the reduction of pain intensity in the short term (two studies: [15,68] $n = 105$, SMD = 0.17, 95% CI −0.22 to 0.55, $Z = 0.84$, $P = 0.4$, heterogeneity: Q-value = 1.7, $P = 0.2$; Figure 9).

Discussion

The systematic review consisted of 12 studies, 8 of which were used to perform the meta-analysis [15,58,61–64,68]. Remarkably, whole studies used for the meta-analysis had acceptable methodological quality according to the PEDro scale.

This is the first meta-analysis of RCTs that has compared GA with other forms of exercise, including GEXP, in CNSLBP patients (more than three months) with associated psychological and disability variables. To better focus on this section, the variables of the main analysis have been discussed.

This meta-analysis shows the current evidence for the effectiveness of GA in CNSLBP patients compared with other forms of exercise, including a CG (waiting list or usual physiotherapy) and GEXP. GA has been demonstrated to be slightly better (i.e., a small effect size) for decreasing disability than the CG. GEXP was shown to be better at reducing catastrophizing and disability than GA. Although these studies found statistically significant effects, the effect of the interventions was small, with an effect size that ranged from 0.29 to 0.53; the largest effect occurred when GA was compared with CG for disability at long-term follow-up. This point will be discussed later.

The variables used for performing the meta-analysis, taking into account the previously mentioned groups,
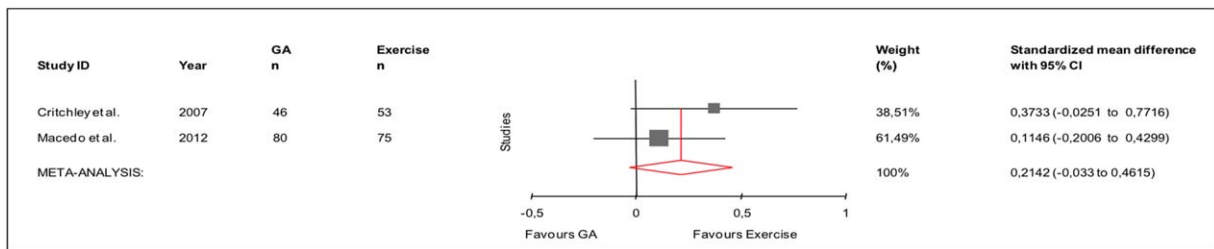
Figure 3 Forest plots of GA vs other forms of exercise effect on quality of life in the long-term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

were as follows: disability, pain, catastrophizing, and quality of life.

Graded Activity vs Other Forms of Exercise

When comparing GA with other forms of exercise, no significant differences were found in the short, intermediate and long term for disability and pain. Neither did other forms of exercise appear to be superior for improving the quality of life in the long term. These results are consistent with numerous reports in the liter-

ature which state that any form of exercise shows the same effectiveness compared with others [14,69]. It is known that exercise itself is beneficial for the treatment of chronic pain, improving all of the factors related to pain in general [70].

Graded Activity vs Control Group

The results of the meta-analysis are in favor of GA in the short, intermediate and long term if the focus is on variables of pain and disability, with the exception of
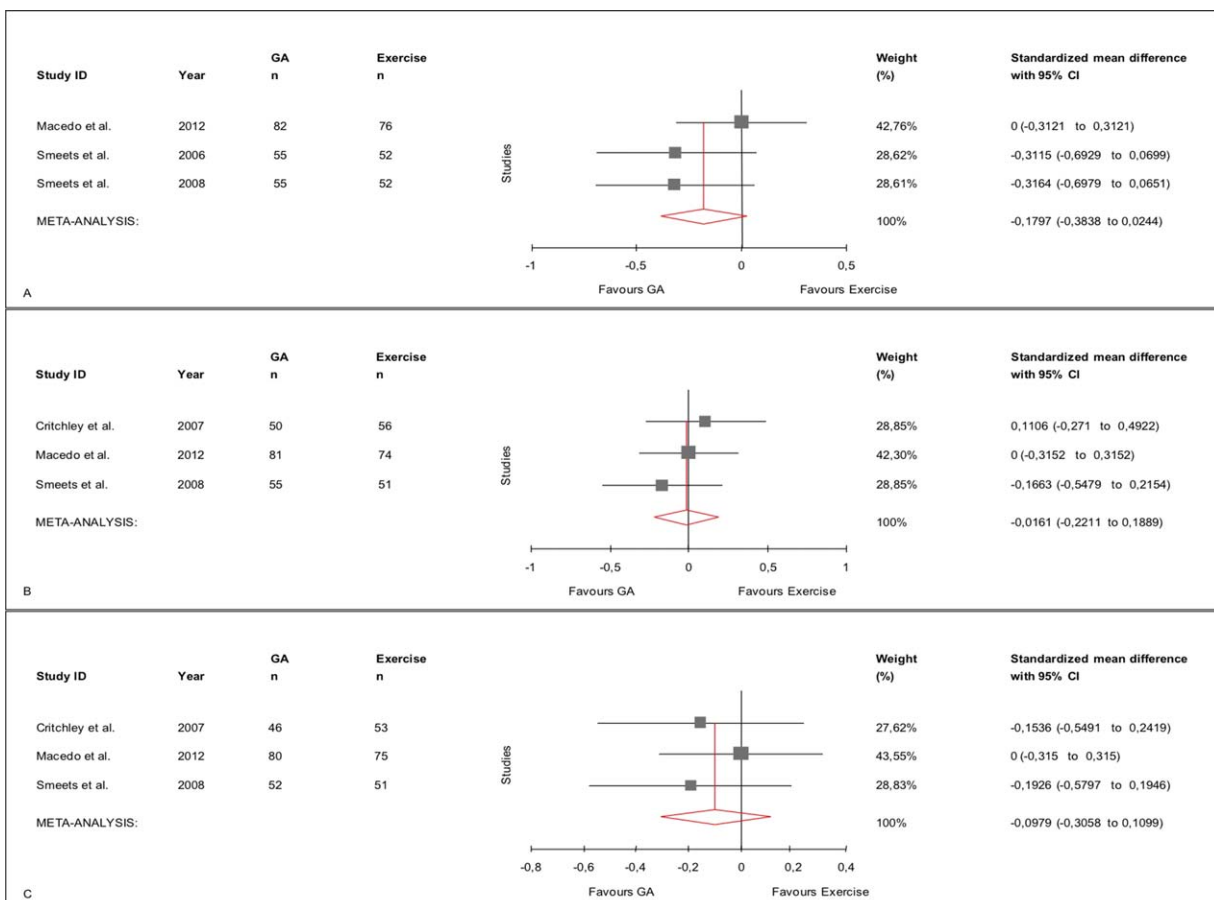

Figure 4 Forest plots of GA vs other forms of exercise effect on pain intensity. (A): short term; (B): intermediate term; (C): long term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
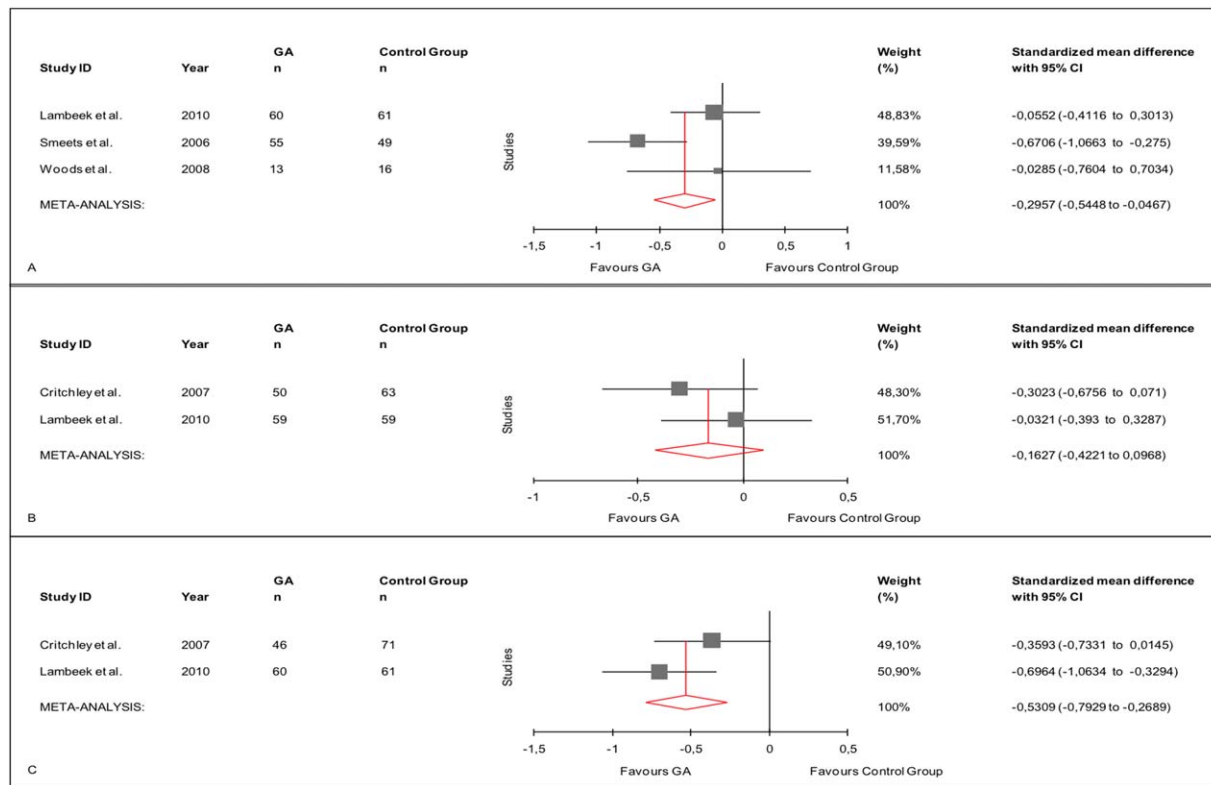
11

Figure 5 Forest plots of GA vs CG effect on disability. (A): short term; (B): intermediate term; (C): long term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

pain in the intermediate term, which are favorable toward the CG. It should be considered that one [68] of the three studies included in short-term pain [60,62,68], was in favor of the CG. In our opinion, this could be because GA is a gradual type of exercise and evidence can be found to demonstrate that exercise shows clear improvements in long-term pain, which is not as obvious in the short and intermediate term [58,60,71,72]. In addition, we observed two clearly significant results in disability, indicating that GA appears to be better in the short or long term, especially the latter, than in the CG. However, the reasons why there are no statistically significant differences in the intermediate term exceed our knowledge. Probably, the short-term score would respond to effects of being observed (Hawthorne effect) [73] through exercise therapy. Conversely, GA therapy needs a learning process, and it is possible that long-term effects could be explained by this process [74]. In this section, the main differences focused on disability but not on pain; this may be because GA focuses on progressively improving the patient's activities without focusing on pain in particular. Therefore, it seems logical that the largest effects would occur in the functionality [18]. In this section, the main differences focused on disability but not on pain; this may be because GA focuses on progressively improving the patient's activities and does not focus on pain in partic-

ular. Therefore, it seems logical that the largest effects would occur in functionality [18].

Graded Activity vs Graded Exposure

Our results clearly show that GEXP is a better short-term option than GA in disability, pain, and catastrophizing. However, these results were based on two RCTs (one acceptable quality RCT and one good quality RCT) that both showed small effect sizes and wide CI; these results should be considered with caution. In our view, a possible explanation for this is that GEXP is focused on the activities that expose the patient's fear. This is likely to reduce their level of fear-avoidance regarding exercise, and therefore, have a catastrophic impact on disability, due to the fact that disability is mediated by the reduction of catastrophizing [75]. Moreover, fear exposure could be a determinant of this. These data are consistent with limited literature that supports success in the reduction of pain, catastrophizing, and disability of GEXP compared with graded exercise [59,66,67]. However, these data differ from the results of Macedo et al. [14], who suggested that the GA is as effective in the short and intermediate term as GEXP or minimal intervention (CG). This may be because their meta-analysis included quasiexperimental studies as well as patients with acute and subacute low back pain (fewer
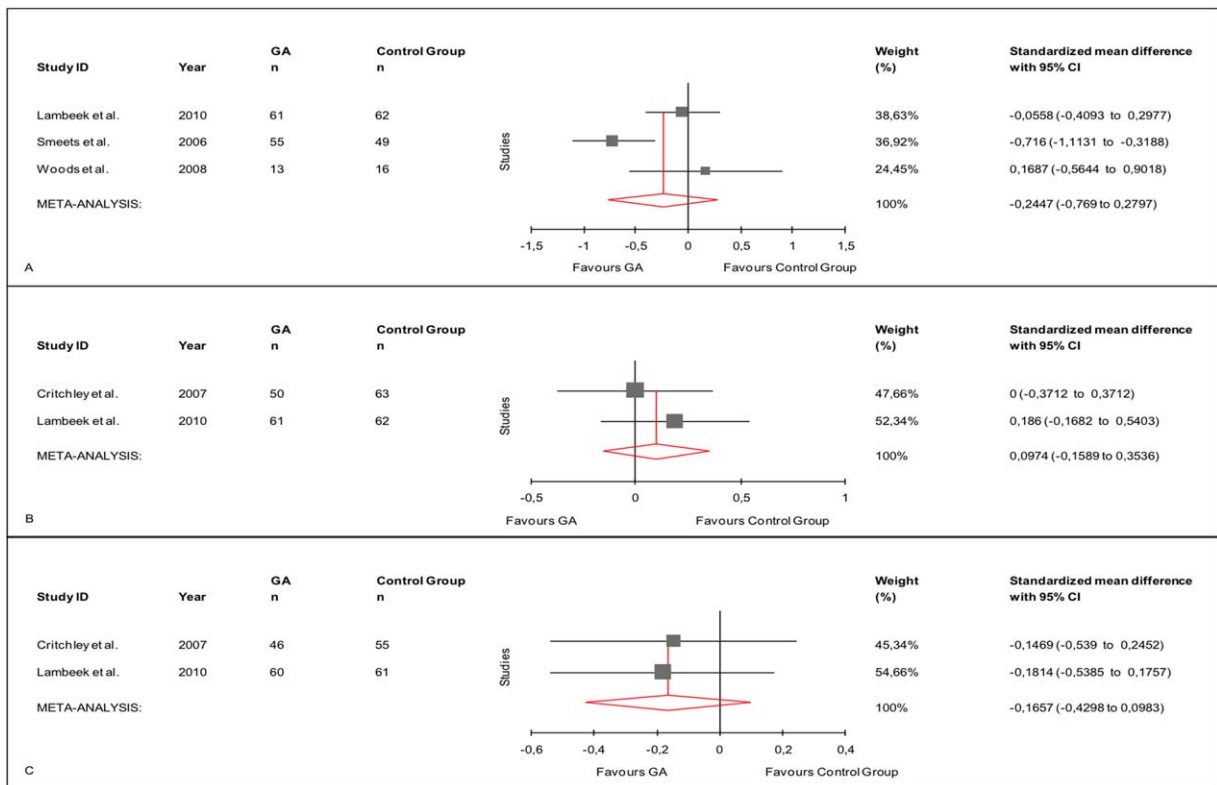
Figure 6 Forest plots of GA vs CG effect on pain intensity. (A): short term; (B): intermediate term; (C): long term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
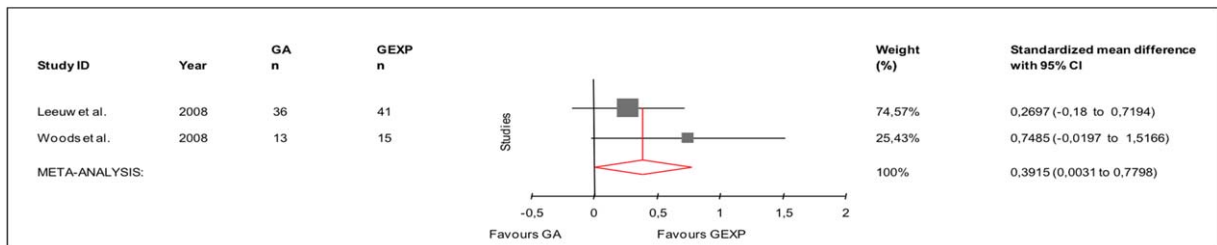

Figure 7 Forest plots of GA vs GEXP effect on disability in the short term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
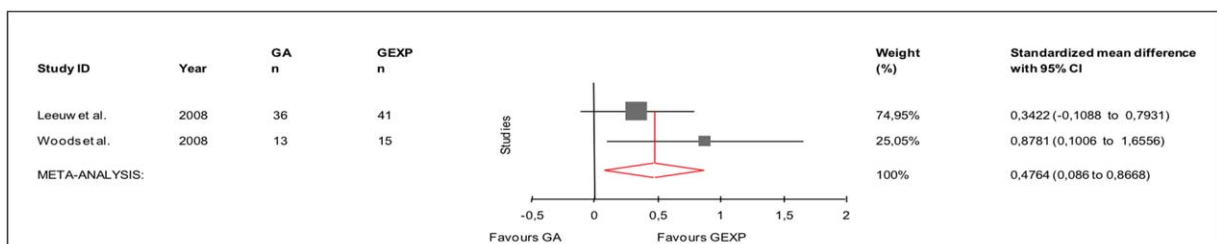

Figure 8 Forest plots of GA vs GEXP effect on pain catastrophizing in the short term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
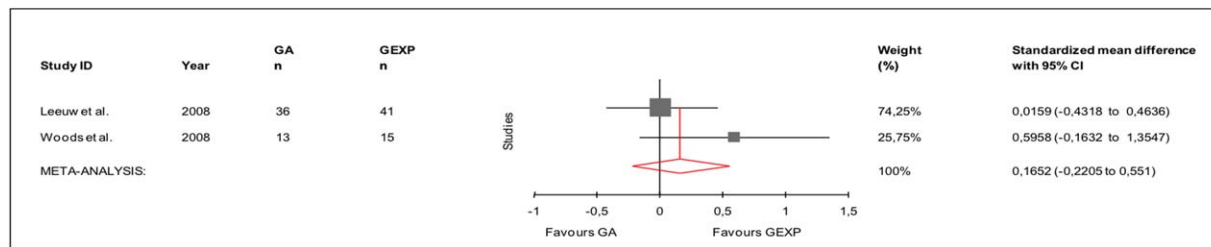
Figure 9 Forest plots of GA vs GEXP effect on pain intensity in the short term. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

than 3 months of symptoms), while we chose only RCTs of CNSLBP. Another difference was that 3 months was considered the boundary between the short and intermediate terms. Later, Van der Giessen et al. [18] supported the results of Macedo et al. [14] when they stated that there is insufficient evidence to assert the superiority of GA vs usual care.

Data obtained from our study are consistent with other systematic reviews and meta-analyses in reference to the suggestion that exercise has a positive effect on reducing pain and disability in particular [76,77]. In our knowledge, there not is other systematic review comparing GA with GEXP in other musculoskeletal conditions different to low back pain.

Implications for Clinical Practice

In clinical practice, it is interesting to consider GA and/or GEXP in subjects suffering from CNSLBP. However, we cannot confirm that GA and/or GEXP are superior to other exercise treatments for patients with CNSLBP. In terms of pain, again we have considered GA and/or GEXP interventions preferable, because there is evidence, albeit limited, that indicates the greater economic profitability of these interventions compared with other forms of physical activity [64]. Therefore, GA and/or GEXP would be more highly recommended in CNSLBP patients because it reduces disability, improves quality of life and decreases the necessary periods of sick leave [78] when compared with the CG. In practice, aggressive patient exposure to fear should be avoided; however, for the treatment of phobias, it is essential that the patient face his own fear. In any case, there are two main facts to be highlighted: on the one hand, the importance of progressive therapies that adhere the patient to treatment and on the other hand, that a patient gradually confronts his fear to find active coping strategies to improve his condition. We have not identified any studies to date that compare the effectiveness of GA and GEXP in the long term (maximum six months). In the short term, GEXP would be recommended over GA to significantly decrease catastrophizing and a nonsignificant trend has been observed in favor of GEXP to improve disability and pain intensity in the short and intermediate term.

Especially important is the fact that the GA has shown Minimal Clinically Important Changes in disability, quality of life, and pain intensity over the long run [58,61,62], while GEXP found clinical changes in catastrophism [15]. All of these changes were obtained for each particular intervention when the follow-up period was compared with baseline.

Limitations

This study has several limitations. First, we have only found a limited number of studies that compare GA with GEXP, and most of these are of poor methodological quality. Second, 2 of the 12 articles selected for systematic review showed poor methodological quality, and the PEDro scale was three points, which was the lowest score. This may distort the results from systematic reviews; however, all studies selected for meta-analysis were of good quality, as recommended by international criteria [79]. Third, all studies lacked blinding of both participants and those responsible for carrying out the treatment, with the exception of the study developed by Woods and Asmundson [68] whose participants were blinded. The use of blinding is very important in the development of high quality trials. Indeed, the nonuse of this system may mean that certain studies do not reach a sufficient level of quality evaluation. While blind outcome assessment was considered achievable [80], it is quite difficult to blind patients or therapists when applying manual therapy or exercise [81]. Although it appears to be complicated, some quality studies have focused on the study of blinding strategies in these fields. Fourth, publication bias could not be assessed in those meta-analyses that included fewer than three studies. Therefore, we cannot determine whether there was publication bias in some of the follow-ups. Fifth, although the results were obtained by weighting the summary estimate of the sample size, different outcomes measures were used to assess the same construct, which might affect the results. Finally, there is a lack of information from gray literature, which has not been examined, such as *Conference Papers Index, Dissertation Abstracts* or *System for Information on Grey Literature in Europe*.

Conclusion

In conclusion, according to the evidence levels described in the methods section, there is limited evidence that GA significantly reduces disability in the short and long term when compared with the CG. In addition, there is strong evidence that GA does not show any differences in pain in the short, intermediate, and long term when compared with the CG. Moreover, there is strong evidence that GA does not lead to differences in disability and pain in the short, intermediate, and long term when compared with other exercise. Finally, there are indicative findings that GEXP is better than GA at decreasing catastrophizing in the short term.

Future high quality investigations should be conducted regarding longer follow-ups that compare GA and GEXP with CGs in CNSLBP patients to support these findings.

References

1 Krismer M, van Tulder M. Strategies for prevention and management of musculoskeletal conditions. Low back pain (non-specific). Best Pract Res Clin Rheumatol 2007;21:77–91.

2 Hoy D, Bain C, Williams G, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012;64:2028–37.

3 Ryan CG, Grant PMM, Dall PM, et al. Individuals with chronic low back pain have a lower level, and an altered pattern, of physical activity compared with matched controls: An observational study. Aust J Physiother 2009;55:53–8.

4 Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: A state of the art. Pain 2000;85:317–32.

5 Wideman TH, Sullivan MJL. Differential predictors of the long-term levels of pain intensity, work disability, healthcare use, and medication use in a sample of workers' compensation claimants. Pain 2011;152:376–83.

6 Dubois J-D, Abboud J, St-Pierre C, Piché M, Descarreaux M. Neuromuscular adaptations predict functional disability independently of clinical pain and psychological factors in patients with chronic non-specific low back pain. J Electromyogr Kinesiol 2014;24:550–7.

7 Caporaso F, Pulkovski N, Sprott H, Mannion AF. How well do observed functional limitations explain the variance in Roland Morris scores in patients with chronic non-specific low back pain undergoing physiotherapy?. Eur Spine J 2012;21(Suppl 2):S187–95.

8 Jensen OK, Nielsen CV, Stengaard-Pedersen K. One-year prognosis in sick-listed low back pain patients with and without radiculopathy. Prognostic factors influencing pain and disability. Spine J 2010;10:659–75.

9 Hodselmans AP, Dijkstra PU, Geertzen JHB, Schiphorst Preuper HR, van der Schans CP. Determinants of change in perceived disability of patients with non-specific chronic low back pain. J Rehabil Med 2010;42:630–5.

10 Rossignol M, Poitras S, Dionne C, et al. An interdisciplinary guideline development process: The Clinic on Low-back pain in Interdisciplinary Practice (CLIP) low-back pain guidelines. Implement Sci 2007;2:36.

11 Danneels LA, Cools AM, Vanderstraeten GG, et al. The effects of three different training modalities on the cross-sectional area of the paravertebral muscles. Scand J Med Sci Sports 2001;11:335–41.

12 Mannion AF, Müntener M, Taimela S, Dvorak J. Comparison of three active therapies for chronic low back pain: Results of a randomized clinical trial with one-year follow-up. Rheumatology (Oxford) 2001;40:772–8.

13 Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev 2005;(3):CD000335.

14 Macedo LG, Smeets RJEM Maher CG, Latimer J, McAuley JH. Graded activity and graded exposure for persistent nonspecific low back pain: A systematic review. Phys Ther 2010;90:860–79.

15 Leeuw, M, Goossens MEJB, van Breukelen GJP, et al. Exposure in vivo versus operant graded activity in chronic low back pain patients: Results of a randomized controlled trial. Pain 2008;138:192–207.

16 George SZ, Zeppieri G. Physical therapy utilization of graded exposure for patients with low back pain. J Orthop Sports Phys Ther 2009;39:496–505.

17 Lindström I, Ohlund C, Eek C, et al. The effect of graded activity on patients with subacute low back pain: A randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992;72:279–90. discussion 291–3.

18 Van der Giessen RN, Speksnijder CM, Helders PJM. The effectiveness of graded activity in patients with

non-specific low-back pain: A systematic review. Disabil Rehabil 2012;34:1070–6.

19 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int J Surg 2009; 8:6.

20 Stone PW. Popping the (PICO) question in research and evidence-based practice. Appl Nurs Res 2002; 15:197–8.

21 Deyo RA, Weinstein JN. Low back pain. N Engl J Med 2001;344:363–70.

22 Ostelo RWJG, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: Towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90–4.

23 Downie WW, Leatham PA, Rhind VM, et al. Studies with pain rating scales. Ann Rheum Dis 1978;37: 378–81.

24 Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine (Phila Pa 1976) 2005;30:1331–4.

25 Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975;1:277–99.

26 Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30:191–7.

27 Roland M, Morris R. A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983;8:141–4.

28 Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec back pain disability scale. Measurement properties. Spine (Phila Pa 1976) 1995;20:341–52.

29 Pollard CA. Preliminary validity study of the pain disability index. Percept Mot Skills 1984;59:974.

30 Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7:524–32.

31 George SZ, Valencia C, Beneciuk JM. A psychometric investigation of fear-avoidance model measures in patients with chronic low back pain. J Orthop Sports Phys Ther 2010;40:197–205.

32 Vlaeyen JW, Geurts SM, Kole-Snijders AM, et al. What do chronic pain patients think of their pain?

Towards a pain cognition questionnaire. Br J Clin Psychol 1990;29:383–94.

33 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.

34 Soer R, Reneman MF, Speijer BLGN, Coppes MH, Vroomen PCAJ. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. Spine J 2012;12:1035–9.

35 Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF36 health survey questionnaire: An outcome measure suitable for routine use within the NHS? BMJ 1993;306:1440–4.

36 Copay AG, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: A choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 2008;8:968–74.

37 Tait RC, Chibnall JT, Krause S. The Pain Disability Index: Psychometric properties. Pain. 1990;40: 171–82.

38 Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain 1983;16:87–101.

39 Katz J, Melzack R. Measurement of pain. Surg Clin North Am 1999;79:231–52.

40 Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002;82:8–24.

41 Reneman MF, Jorritsma W, Schellekens JMH, Göeken LNH. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil 2002;12:119–29.

42 Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. Pain 1986;27:117–26.

43 Osman A, Barrios FX, Gutierrez PM, et al. The Pain Catastrophizing Scale: Further psychometric evaluation with adult samples. J Behav Med 2000;23:351–65.

44 Obradovic M, Lal A, Liedgens H. Validity and responsiveness of EuroQol-5 dimension (EQ-5D) versus Short Form-6 dimension (SF-6D) questionnaire in chronic pain. Health Qual Life Outcomes 2013;11:110

45 Solberg TK, Olsen J-A, Ingebrigtsen T, Hofoss D, Nygaard OP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. Eur Spine J 2005;14:1000–7.

46 Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine (Phila Pa 1976) 2009;34:1929–41.

47 Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609–13.

48 Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration; 2011. Available at: www.cochrane-handbook.org.

49 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.

50 Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998;51:1235–41.

51 Tooth L, Bennett S, McCluskey A, et al. Appraising the quality of randomized controlled trials: Inter-rater reliability for the OTseeker evidence database. J Eval Clin Pract 2005;11:547–55.

52 Bhogal SK, Teasell RW, Foley NC, Speechley MR. The PEDro scale provides a more comprehensive measure of methodological quality than the Jadad scale in stroke rehabilitation literature. J Clin Epidemiol 2005;58:668–73.

53 Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:10.

54 Van Peppen RPS, Kwakkel G, Wood-Dauphinee S, et al. The impact of physical therapy on functional outcomes after stroke: What's the evidence? Clin Rehabil 2004;18:833–62.

55 Bax L, Yu L-M, Ikeda N, Tsuruta H, Moons KGM. Development and validation of MIX: Comprehensive free software for meta-analysis of causal research data. BMC Med Res Methodol 2006;6:50.

56 Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sport Exerc 2009;41:3–13.

57 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088–101.

58 Critchley DJ, Ratcliffe J, Noonan S, Jones RH, Hurley MV. Effectiveness and cost-effectiveness of three types of physiotherapy used to reduce chronic low back pain disability: A pragmatic randomized trial with economic evaluation. Spine (Phila Pa 1976) 2007;32:1474–81.

59 De Jong JR, Vlaeyen JWS, Onghena P, et al. Fear of movement/(re)injury in chronic low back pain: Education or exposure in vivo as mediator to fear reduction? Clin J Pain 2005;21:9–17. discussion 69–72.

60 Lambeek LC, van Mechelen W, Knol DL, Loisel P, Anema JR. Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. BMJ 2010;340:c1035.

61 Macedo LG, Latimer J, Maher CG, et al. Effect of motor control exercises versus graded activity in patients with chronic nonspecific low back pain: A randomized controlled trial. Phys Ther 2012;92: 363–77.

62 Smeets RJEM, Vlaeyen JWS, Kester ADM, Knottnerus JA. Reduction of pain catastrophizing mediates the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. J Pain 2006;7:261–71.

63 Smeets RJEM, Vlaeyen JWS, Hidding A, et al. Chronic low back pain: Physical training, graded activity with problem solving training, or both? The one-year post-treatment results of a randomized controlled trial. Pain 2008;134:263–76.

64 Smeets RJ, Severens JL, Beelen S, Vlaeyen JW, Knottnerus JA. More is not always better: Cost-effectiveness analysis of combined, single behavioral and single physical rehabilitation programs for chronic low back pain. Eur J Pain 2009;13:71–81.

65 Van der Roer N, van Tulder M, Barendse J, et al. Intensive group training protocol versus guideline physiotherapy for patients with chronic low back pain: A randomised controlled trial. Eur Spine J 2008;17:1193–200.

66 Vlaeyen JW, de Jong J, Geilen M, Heuts PH, van Breukelen G. Graded exposure in vivo in the treatment of pain-related fear: A replicated single-case experimental design in four patients with chronic low back pain. Behav Res Ther 2001;39:151–66.

67 Vlaeyen JWS, De Jong J, Geilen M, Heuts PHTG, Van Breukelen G. The treatment of fear of movement/(re)injury in chronic low back pain: Further evidence on the effectiveness of exposure in vivo. Clin J Pain 2002;18:251–61.

68 Woods MP, Asmundson GJG. Evaluating the efficacy of graded in vivo exposure for the treatment of fear in patients with chronic back pain: A randomized controlled clinical trial. Pain 2008;136:271–80.

69 Wang X-Q, Zheng J-J, Yu Z-W, et al. A meta-analysis of core stability exercise versus general exercise for chronic low back pain. PLoS One 2012; 7:e52082.

70 Daenen L, Varkey E, Kellmann M, Nijs J. Exercise, not to Exercise or how to Exercise in Patients with Chronic Pain? Applying Science to Practice. Clin J Pain 2015;31:108–14.

71 Fersum KV, Dankaerts W, O'Sullivan PB, et al. 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. Br J Sports Med 2010;44:1054–62.

72 Costa LOP, Maher CG, Latimer J, et al. Motor control exercise for chronic low back pain: A randomized placebo-controlled trial. Phys Ther 2009;89: 1275–86.

73 Adair JG. The Hawthorne effect: A reconsideration of the methodological artifact. J Appl Psychol 1984; 69:334–45.

74 Patten AR, Sickmann H, Hryciw BN, et al. Long-term exercise is needed to enhance synaptic plasticity in the hippocampus. Learn Mem 2013;20: 642–7.

75 Vlaeyen JWS, Morley S. Cognitive-behavioral treatments for chronic pain: What works for whom? Clin J Pain 2005;21:1–8.

76 Brumitt J, Matheson JW, Meira EP. Core stabilization exercise prescription, part 2: A systematic review of motor control and general (global) exercise rehabilitation approaches for patients with low back pain. Sports Health 2013;5:510–3.

77 Byström MG, Rasmussen-Barr E, Grooten WJA. Motor control exercises reduces pain and disability in chronic and recurrent low back pain: A meta-analysis. Spine (Phila Pa 1976) 2013;38:E350–8.

78 Lambeek LC, Bosmans JE, Van Royen BJ, et al. Effect of integrated care for sick listed patients with chronic low back pain: Economic evaluation alongside a randomised controlled trial. BMJ 2010;341:c6414.

79 Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323:42–6.

80 Minns Lowe CJ, Wilson MS, Sackley CM, Barker KL. Blind outcome assessment: The development and use of procedures to maintain and describe blinding in a pragmatic physiotherapy rehabilitation trial. Clin Rehabil 2011;25:264–74.

81 Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. Control Clin Trials 1987;8:343–53.