Authors:

Susan Armijo-Olivo, PhD, MScPT, BScPT, PT
Jorge Fuentes, PhD, MScRS, BScPT, PT
Bruno R. da Costa, PhD, MScPT, BScPT, PT
Humam Saltaji, DDS, MSc(Ortho)
Christine Ha, BSc
Greta G. Cummings, PhD, RN, FCAHS, FAAN

Affiliations:

From the Department of Physical Therapy, University of Alberta, Edmonton, Alberta, Canada (SA-O, JF); Catholic University of Maule, Department of Physical Therapy, Talca, Chile (JF); Institute of Primary Health Care, Universitat Bern, Gesellschaftstrasse, Bern, Switzerland (BRdC); and School of Dentistry (HS), Rehabilitation Research Center (CH), and Edmonton Clinic Health Academy (GGC), University of Alberta, Edmonton, Alberta, Canada.

Correspondence:

All correspondence and requests for reprints should be addressed to: Susan Armijo-Olivo, PhD, MScPT, BScPT, PT, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada 3-62 Corbett Hall, University of Alberta, Edmonton, Canada T6G 2G4.

0894-9115/16/0000-0000
American Journal of Physical Medicine & Rehabilitation

DOI: 10.1097/PHM.0000000000000521

ORIGINAL RESEARCH ARTICLE

Blinding in Physical Therapy Trials and Its Association with Treatment Effects

A Meta-epidemiological Study

ABSTRACT

Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG: Blinding in physical therapy trials and its association with treatment effects: a meta-epidemiological study. Am J Phys Med Rehabil 2016;00:00–00.

Objective: The aim of this study was to examine whether blinding of participants, assessors, health providers, and statisticians have an effect on treatment effect estimates in physical therapy (PT) trials.

Design: This was a meta-epidemiological study. Randomized controlled trials in PT were identified by searching the Cochrane Database of Systematic Reviews for meta-analyses of PT interventions. Assessments of blinding in PT trials were conducted independently following established guidelines.

Results: Three hundred ninety-three trials and 43 meta-analyses that included 44,622 patients contributed to this study. Only a quarter of the trials were adequately blinded (n = 80; 20%). Most individual components of blinding as well as what they were blinded to were also poorly reported. Although trials with inappropriate blinding of assessors and participants tended to underestimate treatment effects when compared with trials with appropriate blinding of assessors and participants, the difference was not statistically significant (effect size, −0.07; 95% confidence interval, −0.22 to 0.08; effect size, −0.12; 95% confidence interval, −0.30 to 0.06, respectively).

Conclusions: The lack of statistical significance between blinding and effect sizes should not be interpreted as meaning that an impact of blinding on effect size is not present in PT. More empirical evidence in a larger sample is needed to determine which biases are likely to influence reported effect sizes of PT trials and under which conditions.

Key Words: Blinding, Physical Therapy, Randomized Controlled Trials, Risk of Bias

Disclosures:

This research protocol has been approved by the Ethics Board of the University of Alberta (Pro00038172).
This project was funded by the Canadian Institutes of Health Research (CIHR), Alberta Innovates Health Solutions through a knowledge translation initiative grant, the Knowledge Translation (KT) Canada Research Stipend program, and the Physiotherapy Foundation of Canada (PFC) through a B. E. Schnurr Memorial Fund Award. SA-O is supported by the CIHR through a full-time Banting fellowship, the Alberta Innovates Health solution through an incentive award, the STIHR Training Program from KT Canada, and the University of Alberta. GGC holds a Centennial Professorship at the University of Alberta (2013–2020). HS is supported through a Clinician Fellowship Award by Alberta Innovates–Health Solutions, the Honorary Izaak Walton Killam Memorial Scholarship by the University of Alberta, and the Honorary Women and Children's Health Research Institute (WCHRI) Award by the WCHRI.
The funding bodies had no input in the design, collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.
Financial disclosure statements have been obtained, and no conflicts of interest have been reported by the authors or by any individuals in control of the content of this article.

Editor's Note:

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ajpmr.com).

Blinding has been acknowledged as an important criterion for internal validity of randomized controlled trials (RCTs). The purpose of blinding is to prevent both performance bias associated with patients' and research team's expectations and detection bias to prevent bias by assessors.[1,2] Because the term *blinding* is used to describe blinding for several components of a trial (i.e., subjects (participants), health care providers (therapists), data collectors, outcome assessors, or data analysts), various types of blinding can be used in a trial. In addition, several blinding definitions for a trial can be done based on how many components of the trial are blinded (e.g., single, double, or triple blinded). However, these denominations have generated confusion because they actually do not specify who is blinded and mean different things to different people.[3]

Several meta-epidemiological studies investigating associations between trial characteristics and treatment effects have found that trials classified to be not "double blinded" tended to overestimate treatment effects.[4–8] Savovic et al.[8] and Wood et al.,[6] for example, found that the overestimation in trials considered as "not double blinded" was even more pronounced for trials using subjective outcomes such as patient-reported outcomes (pain, disability, quality of life) by 23 to 25%. Nevertheless, other meta-epidemiological studies found no significant association between reported blinding and treatment effects[9–14] or an underestimation of treatment effects of trials with poor blinding.[15]

Overestimates or underestimates of treatment effects, or bias, at the trial level, can lead to biased or inaccurate results and conclusions in systematic reviews and meta-analyses (MAs).[6,16–19] These factors can ultimately have repercussions on clinical decision making and quality of patient care because different assessments could lead to different decisions for clinical practice.

Blinding in physical therapy (PT) and rehabilitation trials is particularly laborious when compared with other fields of medicine.[20,21] The most commonly used control groups in PT trials include interventions as simple as nonparticipation (waiting list), usual care, or another intervention. These control interventions are associated with nonspecific physiologic effects that can affect the active intervention, especially when patients and care providers are not blinded to interventions.[21] Because of the nature of PT interventions (e.g., exercises, devices, manual therapy), blinding for therapists and patients may be challenging and even impossible.[22] Indeed, blinding for health care providers, patients, and outcome assessors is less frequently reported in trials involving nonpharmacological interventions.[23] Although progress in reporting blinding in PT and rehabilitation trials has been certainly documented in the last decade,[21] this is still incomplete and does not fulfill current recommendations.[21]

In addition, the type of outcome is crucial when evaluating blinding. Objective and automated outcomes (i.e., mortality, laboratory results, and administrative data) may be less prone to assessment bias than subjective outcomes such as pain, which is a self-reported measure and commonly used in PT.[6] Therefore, trials evaluating effects of PT treatments such as on mortality or hospitalizations (automated data obtained from administrative hospital databases) may be less susceptible to bias than trials evaluating effects of PT treatments on pain or self-reported measures. However, this theoretical thinking needs to be proven based on empirical evidence in the PT field.

Most of the empirical evidence regarding associations between blinding and treatment effect estimates comes from RCTs in medicine and is based

mainly on dichotomous outcomes and looking only at limited aspects of blinding (i.e., double blinding).[6,7,15] The closest meta-epidemiological study to PT area was performed in trials of osteoarthritis investigating several types of treatments such as drug trials and other nonpharmacological interventions.[5] No such comparative studies using continuous outcomes have been conducted in other health areas such as the allied health professions, including PT.

According to recent meta-epidemiological studies,[2,8] the effect of blinding is highly unpredictable, and separate analyses of blinding effects in individual trials and MAs for performance (blinding of participants and personnel) and detection biases (blinding of assessors) are needed. Therefore, empirical evidence is needed in PT to determine the extent to which blinding and forms of blinding affect treatment effect estimates. This information will support development of future guidelines for designing, conducting, implementing PT trials, and assessing quality/risk of bias of PT trials in systematic reviews and MAs and ultimately the strength of evidence for application in clinical decision making.

Our objectives were to (1) describe use of different forms of blinding; (2) determine appropriateness of its reporting in PT RCTs; (3) examine whether blinding of participants, assessors, health providers, and statistician has an effect on treatment estimates in PT trials; and (4) determine if these effects differ depending on characteristics of the MAs analyzed, such as magnitude of effect size (ES), MA heterogeneity, type of outcome (subjective or objective), and whether the MA involves the musculoskeletal area.

METHODS

This study is part of a large meta-epidemiological project investigating the association between biases and treatment effect estimates in PT trials. The protocol for this meta-epidemiological study was published previously.[24]

Study Selection

The Cochrane Database of Systematic Reviews was systematically searched from 2005 to May 25, 2011, for MAs of PT interventions using the following words: PT or physiotherapy, rehabilitation, exercise, electrophysical agents, acupuncture, massage, transcutaneous electrical stimulation, interferential current, ultrasound, stretching, chest therapy, pulmonary rehabilitation, manipulative therapy, mobilization, and related terms. Details of the search strategy can be found elsewhere.[25] Meta-analyses and their RCTs were included if (1) the MA included at least 3 RCTs comparing at least 2 interventions, with at least one of the interventions being currently or potentially part of PT practice according to the World Confederation for Physical Therapy[26]; (2) the outcome of interest in the MA (main outcome or with the largest number of trials) was continuous. We decided to use continuous outcomes for this project for 2 reasons: (1) to extend the knowledge base of meta-epidemiological studies to continuous outcomes because most available evidence from these studies has been modeled on dichotomous outcomes such as all-cause mortality and presence of events.[4,6–10,12,13,15,16,27–29]; and (2) continuous outcomes, such as pain, disability, and range of motion, are commonly used in PT trials, which makes our information more applicable to practice.

A unique code generated by the Reference Manager bibliographic program was assigned to each MA and trial that met the inclusion criteria. This code was used to randomly select studies to be analyzed and also to randomize the order of evaluation. The first author (SA-O) randomly selected each MA to be included and accompanying trials by drawing the code of the selected MA first and then the selected trial from an opaque envelope. This process ensured that the researcher had no influence on the studies selected or the order of evaluation.

Assessment of Blinding Domains

We evaluated specific forms of blinding for each trial component such as participants (i.e., blinding of individuals who are randomly assigned to interventions under evaluation), assessors (blinding of individuals who collect data for trial outcomes), therapists (blinding of clinicians, i.e., physiotherapists or other care providers who care for participants during the trial), statistician (blinding of individuals who conduct trial analyses), and investigators (blinding of individuals who are nominated as trial principal investigators), and overall appropriateness of blinding (whether the appropriate component of blinding [participants, assessors, or irrelevant] based on the main outcome of interest) was used as suggested in the most commonly used tools used in PT (i.e., Delphi List, PEDro, Maastricht, Maastricht-Amsterdam List, Bizzini, van Tulder and Jadad tools).[30] A 3-point scale (yes, no, unclear) was used to assess items, and guidelines for scoring were extracted from the guidelines of these tools.

Furthermore, we evaluated whether each individual component of a trial would be blinded to (1) study hypothesis, (2) details of interventions, (3) random assignment, (4) outcome measures, and (5) outcome analysis. In addition, we evaluated whether

blinding was successful (whether the blinding was maintained in the trial and the participants/assessors/personnel did not guess the random allocation) as suggested by Friedberg et al.[31] We based the assessment of blinding on the main outcome described in the MAs.

Data Extraction of Treatment Estimates and Trial Characteristics

Two independent reviewers extracted (any of these reviewers: SA-O, JF, HS, CH, AC, DP) data from each trial based on each MA. Data on means, SDs, sample sizes, type of interventions (and their details such as intensity, frequency, and dosage), information related to type of outcome (i.e., objective, subjective), funding source, publication year, design characteristics, and statistical analysis were extracted. The outcome of interest selected for each MA in the Cochrane review that met the eligibility criteria was chosen as the primary outcome for analysis. If not clearly specified, the outcome was determined according to the MA that contained the largest number of trials in the review.

Characteristics of the Reviewers' Panel

Six reviewers with experience in different areas of health sciences research comprised the review panel for this study. Characteristics of the review panel can be found elsewhere or upon request.[25,32,33]

Analysis

Data Synthesis

We carefully followed the same methodology to perform meta-epidemiological analysis described in predominant studies in this field.[5–8,12–16,27,34,35] In order to determine whether forms of blinding affect treatment effect estimates, a 2-level analysis was conducted using a meta–meta-analytic approach with a random-effects model to allow for within and between MAs heterogeneity as described by Sterne et al.[27] The first-level analysis (within MA) was as follows: we derived ESs for each trial by dividing the between-group difference in mean values by the pooled SD as described by Cohen.[36] A negative ES indicates a beneficial effect of the experimental intervention. If some required data were unavailable, we used approximations as previously described.[37] Data from each trial were obtained from each MA. We followed the Cochrane reviews to determine the comparison included for analysis (i.e., treatment of interest and control group). In the case of studies appearing in more than 1 review, the study was considered only once in the MA with the fewer number of overall studies. We then calculated 2 pooled ES for each MA: one corresponding to the pooled ES from those studies having the characteristic of interest (e.g., blinding assessment) and the other for those studies that did not (e.g., no or unclear blinding assessment). We used standard random-effects MAs to combine ES across trials and calculated the DerSimonian and Laird estimate of the variance to determine heterogeneity between trials.[5,34] Then, for each MA, we derived the difference between pooled ES estimates from trials with and without the characteristic of interest (e.g., blinding of assessor). A negative difference in ES indicates that trials with the characteristic of interest show a more beneficial effect for the experimental group.

The second-level analysis (between MAs) involved pooling the results of the previous analysis to describe the effect of each trial component across all MAs. The ESs were also combined at this stage using the DerSimonian and Laird random-effects models[38] to allow for between-MA heterogeneity.

Formal tests of interaction between adequate blinding component and estimated treatment benefits were performed separately for each MA based on Z scores using the estimated difference in ES between trials with and without adequate blinding.

We additionally stratified analyses accompanied by interaction test based on Z scores according to prespecified characteristics: treatment benefit in overall MA: small (ES >-0.5) versus large (ES ≤-0.5); between-trial heterogeneity in overall MA (low [$\tau^2 < 0.06$] vs. high [$\tau^2 \geq 0.06$]), nature of outcome (subjective or objective), and if the intervention was classified as musculoskeletal or other PT area. The prespecified cutoff of $\tau^2 = 0.06$ corresponds to a difference between the smallest and largest ES of approximately 1 ES.

In order to evaluate the effect of different types of blinding on treatment effect estimates, we limited analyses to studies describing a true control group (i.e., group receiving no treatment or a waiting list) or placebo intervention and studies in which the direction of expected treatment effect was evident (i.e., standard care vs. standard care plus active intervention; and active intervention 1 plus active intervention 2 vs. active intervention 1). Again, a DerSimonian and Laird random-effects model was used to pool data. Stata statistical software version 12 (StataCorp LP, College Station, TX) was used to perform these analyses.

RESULTS

Selection and Characteristics of MAs and RCTs

The search identified 3901 Cochrane reviews, with 271 reviews potentially relevant to PT. Of

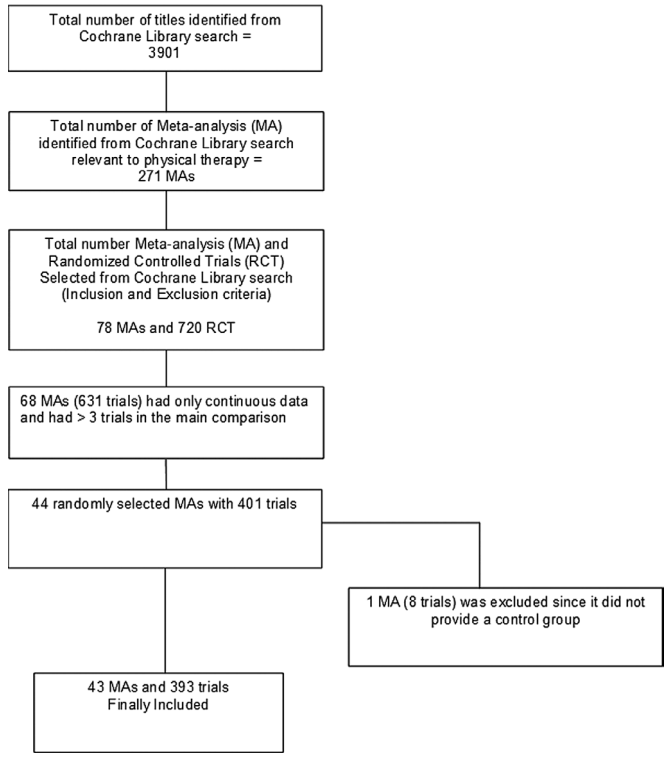

FIGURE 1 *Diagram for identification of studies.*

these, 68 reviews included an MA of at least 3 studies of PT interventions and examined at least 1 continuous outcome. We randomly selected 44 MAs but excluded one[39] because it used follow-up data from the same group rather than a control group for comparison (Fig. 1). Forty-three MAs including 393 trials and analyzing 44,622 patients contributed to this study. Table 1 summarizes the characteristics of the 43 Cochrane reviews. Briefly, reviews were published between 2008 and 2011 and included MAs of the effectiveness of PT interventions for musculoskeletal (22 reviews),[40–48] cardiorespiratory (9 reviews),[49–57] neurological (6 reviews),[58–64] and other PT areas (6 reviews).[64–69] Most trials were parallel-group trials (367; 93.4%) and single-center studies (298; 76%) and had active control interventions (362; 92%). The most common intervention was exercise (n = 282; 71.8%). Remaining trials used a combination of exercise and physical agents, manual therapy, and other treatments such

TABLE 1 Characteristics of the selected MA within PT areas

	Musculoskeletal	Cardiorespiratory	Neurology	Other
Characteristics				
No. of MAs	22	9	6	6
Median year of publication	2009	2010	2009	2009
Total no. of trials included	194	78	52	69
Total no. of patients included	19861	9015	2138	13608
Main intervention				
Exercise	13	7	3	5
Physical agents	1	0	1	0
Acupuncture	2	0	0	0
Manual therapy	1	0	0	0
Other	1	2	2	1
Outcomes				
Clinician-assessed outcome	8	4	6	3
Self-reported outcome	11	4	0	2
Administrative data or automated outcome	3	1	0	1

as respiratory exercises. Supplementary Table S1 (Appendix 1; http://links.lww.com/PHM/A245) lists characteristics of the 43 MAs.

Reporting of Blinding

Details of reporting of blinding by participants, assessors, investigators, therapists, and statisticians and what they were blinded to (hypothesis, details of intervention, random assignment, and analysis) can be found in Table 2. General description of blinding for each component is as follows.

Blinding of Participants

Blinding of participants to random allocation was used in only 31 (7.9%) of 393 trials. The remaining trials either did not blind participants (n = 88; 22.4%) or were unclear whether participants were blinded (n = 272). Participants were reported as blinded to the study hypothesis in 20 trials (5.1%). In addition, only a small number of trials (n = 13; 3.3%) reported that participants were blinded to intervention details, whereas the vast majority of them (n = 379; 96.4%) did not blind subjects to intervention details (n = 313). Only 7% of trials (n = 27) blinded participants to outcome assessment. None of the trials reported that they kept participants unaware of data analysis. Only 4 trials reported that blinding was successful, and 5 reported that it was not. The rest of the trials did not report evaluation of participants' blinding (Table 2).

Blinding of Assessors

Of 393 trials, blinding of outcome assessors to random allocation was used in 95 trials (24%). Assessors were reported blinded to the study hypothesis in only 13 trials (3.3%). In addition, 41 trials (10.4%) reported that assessors were blinded to intervention details. None of the trials reported that they kept assessors unaware of data analysis. Only 6 trials (1.5%) reported that blinding of assessors was successful. The rest of the trials did not report, or it was unclear if they evaluated the success of assessors' blinding (Table 2).

Blinding of Investigators

Of 393 trials, blinding of investigators to random allocation was reported in 13 trials (3.3%). Trials did not report whether investigators were blinded to statistical analysis. In addition, none of the trials reported whether blinding of investigators was successful. Only 2 trials (n = 0.5%) reported to blind principal investigators to outcome measures.

TABLE 2 Type of blinding in PT trials

	Random Allocation			Hypothesis			Intervention Details			Outcome Assessment			Data Analysis			Blinding Successful		
Blinding	Yes	Unclear	No	Yes	Unclear	No	Yes	Unclear	No	Yes	Unclear	No	Yes	Unclear	No	Yes	Unclear	No
Participants	31	272	88	20	367	6	13	67	313	27	182	184	0	393	0	4	384	5
%	8	69	22	5	93	2	3	17	80	7	46	47	0	100	0	1	98	1
Assessors	95	237	61	13	377	3	41	198	154	NA	NA	NA	0	393	0	6	380	7
%	24	60	16	3	96	1	10	50	39	NA	NA	NA	0	100	0	2	97	2
Investigators	13	367	13	NA	NA	NA	NA	NA	NA	2	387	4	0	393	0	0	393	0
%	3	93	3	NA	NA	NA	NA	NA	NA	1	98	1	0	100	0	0	100	0
Therapists	2	0	391	1	384	8	NA	NA	NA	12	368	13	0	393	0	0	393	0
%	1	0	99	0	98	2	NA	NA	NA	3	94	3	0	100	0	0	100	0
Statisticians	15	375	3	0	393	0	0	393	0	0	393	0	NA	NA	NA	0	393	0
%	4	95	1	0	100	0	0	100	0	0	100	0	NA	NA	NA	0	100	0

NA, not applicable.

Remaining trials either did not report that principal investigators were blinded to outcome measures (n = 387; 98.5%) or clearly stated that they did not blind principal investigators to outcome measures (n = 4; 1%). None of the trials in this category reported blinding of study hypothesis and intervention details. However, principal investigators are aware of these features because they are designing and planning the trials (Table 2).

Blinding of Therapists

A minimal percentage of trials (only 2 trials [0.5%]) reported blinding of therapists. Remaining trials did not blind therapists because this was not possible (n = 391; 99.5%). Therapists were reported blinded to the study hypothesis in only 1 trial (0.3%). Only 3% of trials (n = 12) blinded therapists to outcome assessments. None of the trials reported keeping therapists unaware of data analysis. None of the trials reported whether blinding of therapists was successful (Table 2).

Blinding of Statisticians

Blinding of statisticians to random allocation was reported in only 15 trials (3.8%). Remaining trials either did not blind the statistician (n = 3; 0.8%) or were unclear whether statisticians were blinded or not (n = 375; 95.42%). Trials did not report whether statisticians were blinded to study hypothesis and intervention details. None of the trials reported that statisticians were blinded to outcome measures as well. None of the trials reported whether blinding of the statistician was successful (Table 2).

Double Blinding

Of 393 trials, only 10 trials were reported to be double blinded (2.54%). The definition of double blind varied. Some trials defined double blind when both participants and investigators were blinded. Another trial was reported as double blinded when participants, assessors, data managers, and statisticians were blinded. Another trial was stated to be double blinded, but only specified participants as blinded. Remaining trials (n = 7) defined double blinding as participants and assessors being blinded.

Overall Blinding Adequacy

Of 393 trials, 80 were considered as having appropriate blinding (20.4%). Remaining trials either were considered to have no appropriate blinding (n = 42; 10.7%) or were unclear whether they had appropriate blinding (n = 271; 69%).

Blinding of Participants and Treatment Effects in PT Trials

To analyze the effect of blinding of assessors on treatment effects, 9 MAs including 121 trials and analyzing 13,151 patients were used. Figure 2 shows the forest plot of differences in ESs between trials with adequate and inadequate blinding of participants. Although trials with inappropriate blinding of participants tended to underestimate treatment effects when compared with trials with appropriate blinding of assessors, the difference was non–statistically significant (ES, −0.12; 95% confidence interval [CI], −0.30 to 0.06). Results of stratified analyses are displayed in Figure 3. In this

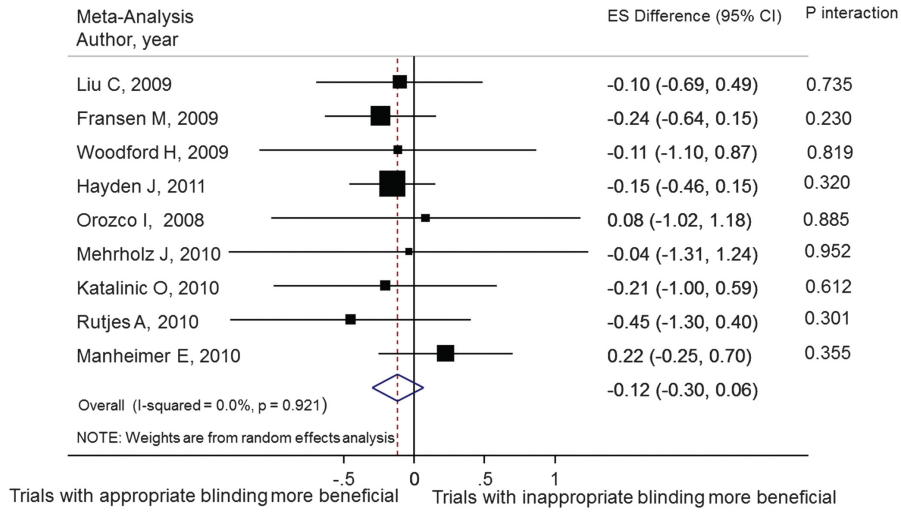

FIGURE 2 *Forest plot of the differences in ESs between trials with and without appropriate blinding of participants.*

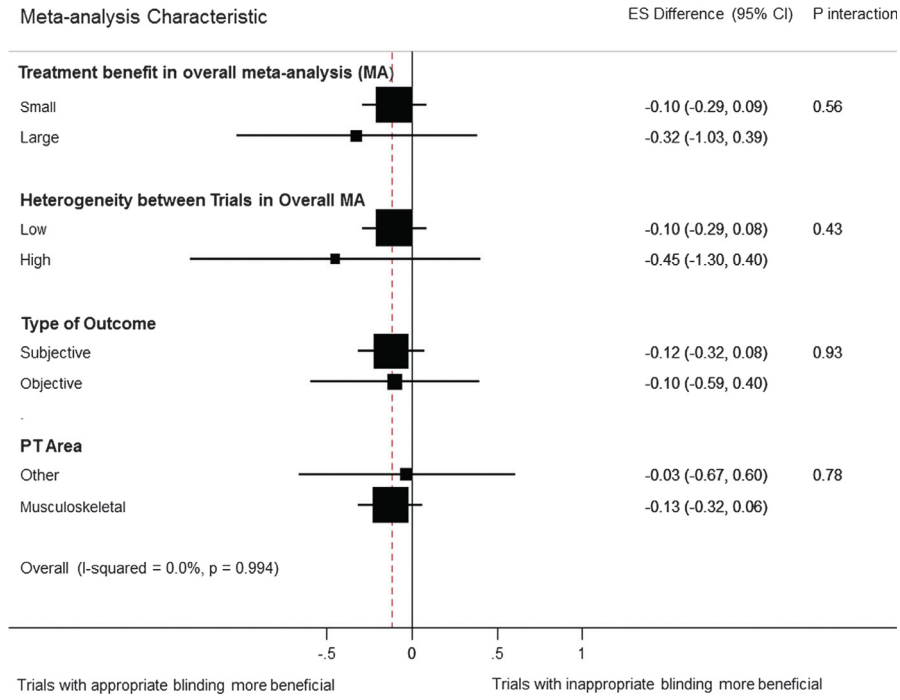

FIGURE 3 *Forest plot of the differences in ESs between trials with and without appropriate blinding of participants stratified by MA characteristics (ES magnitude, heterogeneity, type outcome, and PT area).*

stratified analyses, none of the MA characteristics had a statistically significant interaction.

Blinding of Assessors and Treatment Effects in PT Trials

Seventeen MAs including 165 trials and analyzing 23,316 patients were used to analyze the effect of blinding of assessors on treatment effects. Figure 4 shows the forest plot of the differences in ESs between trials with adequate and inadequate blinding of assessors. Although trials with inappropriate blinding of assessors tended to underestimate treatment effects when compared with trials with appropriate blinding of assessors, the difference

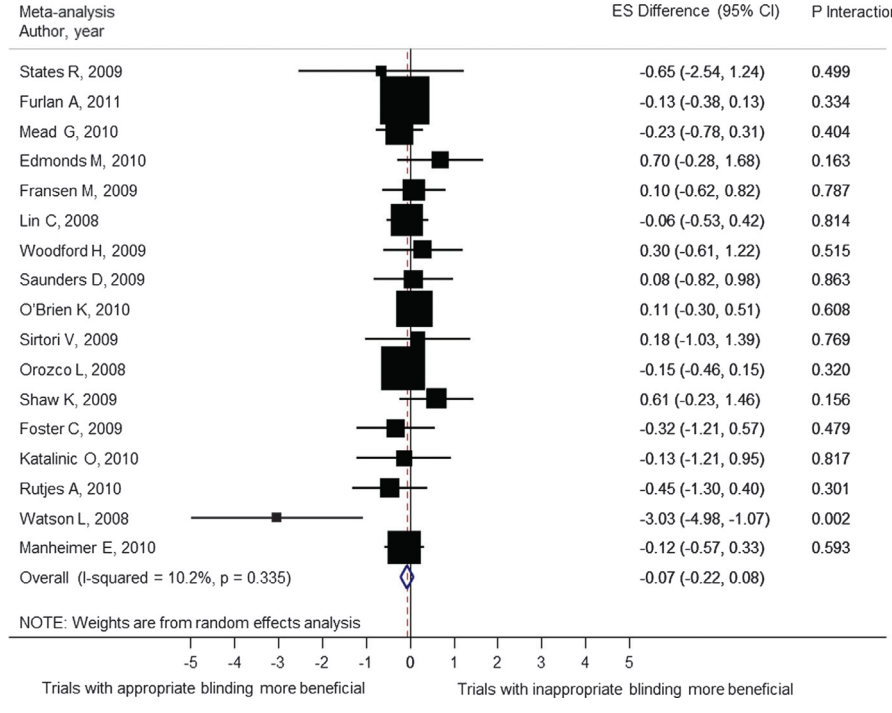

FIGURE 4 *Forest plot of the differences in ESs between trials with and without appropriate blinding of assessors.*

was non–statistically significant (ES, −0.07; 95% CI, −0.22 to 0.08). None of the MA characteristics had a statistically significant interaction.

Blinding of Investigators and Treatment Effects in PT Trials

To analyze the effect of blinding of investigators on treatment effects, 7 MAs including 113 trials and analyzing 10,854 patients contributed to this analysis. Trials with inappropriate blinding of investigators tended to have an overestimated treatment effect when compared with trials with appropriate blinding of investigators, although the difference was non–statistically significant (ES, 0.10; 95% CI, −0.17 to 0.38). None of the MA characteristics had a statistically significant interaction.

Blinding of Statistician and Treatment Effects in PT Trials

Six MAs including 89 trials and analyzing 11,363 patients contributed to the analysis of the effect of blinding of investigators on treatment effects. These results showed that the difference between trials with inappropriate blinding of statistician and trials with appropriate blinding of statistician was non–statistically significant (ES, 0.04; 95% CI, −0.18 to 0.27). None of the MA characteristics had a statistically significant interaction.

Overall Adequacy of Blinding and Treatment Effects in PT Trials

For the purpose of analyzing the overall effect of adequacy of blinding on treatment effects, 16 MAs including 182 trials and analyzing 16,019 patients contributed to this analysis. Although trials with inappropriate overall blinding tended to underestimate treatment effects when compared with trials with appropriate overall blinding, the difference was non–statistically significant (ES, −0.08; 95% CI, −0.28 to 0.12). None of the MA characteristics had a statistically significant interaction.

DISCUSSION

The results of this study showed that reporting and conducting blinding in PT are poor. Only a quarter of trials were adequately blinded. Remaining trials were considered either unclear (69%) or not appropriately blinded (11%). Most individual components of blinding (e.g., participants, assessors, statisticians, and investigators) and what they were blinded to (hypothesis, details of intervention, random assignment, analysis, etc.) were also poorly reported. Results of our meta-epidemiological approach show that trials with inadequate blinding of assessor, participant, and overall appropriateness of blinding tended to underestimate treatment effects, although differences were not statistically significant.

Underreporting and Difficulty of Blinding in PT Trials

Of note, it is important to highlight that a large percentage of trials did not clearly report if blinding was accomplished or not for different components of the trial. The large amount of underreporting of blinding in trial reports has been previously reported, especially for nonpharmacological trials.[20,23] Although reporting of methodological factors in physical medicine, rehabilitation,[21] and PT[70] has improved in recent years,[21] reporting of blinding and specifically blinding of participants has not increased at all, and reporting of blinding of therapists has actually decreased every year according to a recent report.[70] In addition, Moseley et al.[70] found that reporting of blinding of assessors has not increased as much as other methodological items, with only a 3.4% of increase reporting by year.[70] Thus, it seems that adequate reporting of blinding is lacking in rehabilitation, especially in PT trials. Boutron et al.[71] highlighted that lack of blinding and its underreporting could be due to lack of awareness of available and innovative methods for blinding in nonpharmacological trials. Thus, we have compiled a table with possible ways to enhance blinding in the PT and related fields (see Appendix 2; http://links.lww.com/PHM/A246).

In addition, poor reporting has been linked to blinding being harder to perform in nonpharmacological trials such as PT trials.[20] Only 22% of trials assessing rehabilitation, acupuncture, education, and spa therapy were judged as possible to blind patients in an earlier report.[20] Perceived risk of not blinding was more often considered "moderate" or "important" in nonpharmacological trials than pharmacological trials for patients and outcome assessors, thus introducing biases to study results.[20] Although these examples are not exhaustive, it can provide researchers working in PT and rehabilitation with some guidance on how to blind certain components of the trials to avoid performance or detection biases.

Most examples applicable to rehabilitation trials could produce a placebo effect when using physical agents' therapies or when applying acupuncture, although this requires innovative solutions to make it credible. For example, recent systematic reviews highlighted that new types of needles can be used to

produce effective and credible placebos for acupuncture,[72] especially one known as the "Takakura device," which shows promising results to blind both participants and therapists.[73] In addition, some studies have used naive participants to certain PT treatments such as transcutaneous electrical stimulation or interferential current to make a placebo treatment credible.[74] However, when using an attention-control group or other active (supposedly non therapeutic) interventions such as hands-on procedures without adjustments or manipulation, blinding of participants is challenging because the generation of a credible placebo controlled trial in these circumstances is generally not possible or cumbersome.[22] One solution would be to blind participants, assessors, statisticians, or other personnel to trial hypothesis and purposes. In this way, performance biases could be reduced; however, it comes with ethical considerations. Furthermore, when possible, objective outcomes or outcomes that can be measured by an external blinded assessor or panel are used as suggested by the use of the PROBE (Prospective Randomised Open Blinded Endpoint) design.[75] However, this type of design is limited when the main outcome is a patient-reported outcome.

Other possible strategies to decrease the possibility for biases, especially at the therapist level, would be expertise-based RCTs.[76] Therapists participating in trials of competing interventions (e.g., manipulation vs. acupuncture) usually have their own preconceived opinion about which treatment will be more effective and are likely to believe that the type of treatment they practice and have expertise in will be more effective. As a result of unblinding therapists, it is very unlikely that 1 clinician will maintain clinical equipoise when using 2 competing treatments. Generally, they could be biased to perform one treatment better than the other, and they could potentially prescribe differential cointerventions to study groups. Furthermore, if therapists are also data collectors of main outcomes in the trial, they may differentially record data or interpret outcomes depending on whether patients received the form of therapy that they practice. Thus, using expertise-based RCTs would minimize differential-expertise bias by allocating participants to clinicians with expertise in the specific interventions under investigation who are committed to performing either intervention A or intervention B based on their expertise. This randomization would also increase fidelity of the intervention.

It should be noted that when evaluating the influence of blinding on treatment effects, other confounding factors, such as participant expectation/experience and practitioner-participant interaction, may affect the therapeutic effect.[74] In our set of analyzed trials, the majority of included studies failed to report on whether these issues were considered and what precautions were taken. Future RCTs should report more details on how much information was given to participants regarding the interventions and how therapists-participant interactions were managed (limited or encouraged).

Another method proposed to prevent performance bias is using the so-called "Zelen" randomization method. According to a review,[77] the Zelen method was commonly used to limit bias, especially to avoid "resentful demoralization (i.e., disappointment by not getting the alternative treatment, which affects response to treatment and study outcomes) and the Hawthorne effect (i.e., a change in behavior occurring because of trial participation rather than any treatment) generally occurring in conventional RCTs. When using the Zelen method, the patients in the control group are not aware of the presence of an alternative therapy; thus, resentful demoralization could be reduced and could also potentially decrease the dropout rate, reporting of adverse events, and negative results, which has been commonly acknowledged in conventional RCTs.[78] One of the drawbacks of the Zelen design is the presence of crossovers, which will dilute the treatment effect in trials using this type of design. In addition, the Zelen method has been criticized ethically because of lack of consent.

Blinding: Reporting and Conduct

It is possible that authors do not report blinding in their articles, but they do perform adequate blinding in their actual trials. Devereaux et al.[79] found that authors fail to report in their articles blinding status of participants in 26%, health care providers in 64%, data collectors in 84%, outcome assessors in 83%, and data analysts in 96% of them. Nevertheless, trials that were not reported to be blinded actually used blinding in 20%, 65%, 65%, 79%, and 50% of participants, health care providers, data collectors, outcome assessors, and data analysts, respectively.[79] Thus, it is possible that lack of reporting does not necessarily mean poor conduct. However, Devereaux et al.[79] analyzed a subset of studies (n = 98) in the area of internal medicine, and thus, this information might not be applicable to PT and rehabilitation areas.

Because the term *blinding* is used to describe blinding for several components of a trial, a lot of confusion exists in describing who is really blinded. The term *double blinding* may refer to blinding 2 of the components of a trial and any combination of

them, for example, participants and health care providers, investigators and data collectors, and assessors and data analysts. Thus, the definitions of single, double, or triple blinded vary depending on who is making the judgment, which makes the use of these descriptions confusing and ineffective. This was evidenced in the results of our study because several combinations were used to describe double blinding. In order to avoid confusion, the Consolidated Standards of Reporting Trials (CONSORT) Statement recommends discard the terms single, double, or triple blinding. Instead, the CONSORT statement suggested that authors report detailed blinding-related parameters, including who was blinded and how, and explain the mechanisms of blinding and reasons if any key trial collaborator is not blinded.[80] Also, the CONSORT statement recommends avoiding the assessment of the success of blinding because of lack of empirical evidence and validity supporting its practice.[81] In addition to recommendations provided by the CONSORT statement, and based on the results of this study, we recommend that authors of primary RCTs report which components of the trial people were blinded to (e.g., hypothesis, details of intervention, random assignment, outcome assessment, analysis) and the implications of blinding or lack of blinding in study results. In this way, it is easier to evaluate the influence of these biases in the trial's results. For example, if the main outcome of a trial is self-reported and participants were not blinded to random allocation, blinding participants to study hypotheses could minimize performance biases, and thus, results of the study could be viewed as less biased than only having the information of not being blinded. Journal editors and reviewers have an essential role in this aspect. They should make sure that a clear description of each of the components of blinding is performed. This will ensure that the quality of reporting of future trials is improved and that the evaluation of the risk of bias of RCTs in this field is more accurate.

The Relationship Between Blinding and Treatment Effect Estimates

Lack of blinding in RCTs has been associated with increased magnitudes of observed treatment effects.[4–6,8] Some meta-epidemiological studies have found that lack of double blinding may overestimate treatment effects by 9% to 44%[4–6,8] Nevertheless, other meta-epidemiological studies found no significant association or underestimation of treatment effects in trials with inappropriate blinding of patients, therapists, and outcome assessors.[9–11,14,15] The results of the present study are in agreement with these reports. There was no statistically significant difference between trials with and without appropriate blinding of participants, assessors, investigators, or statisticians or overall blinding on treatment effects. Most of the previous reports have wide CIs around the estimates, which does not rule out the possible overestimation or underestimation of effects overall. This is in line with the results of the present study, which is also characterized by wide CIs. In addition, the direction of bias for blinding has been inconsistent. As opposed to the general finding of overestimation of treatment effect for trials of poor quality, this study shows that trials with inappropriate blinding tended to show a decreased effect when compared with trials with adequate blinding. A recent report[12] suggested that results for meta-epidemiological studies investigating evidence of bias and treatment effect estimates varied between meta-epidemiological data sets, individual MAs within data sets, and clinical fields. Thus, the direction of the bias is not consistent for all areas of research, and more research is needed to provide more insights regarding the direction of the biases.[2,8]

In addition, with lack of reporting of blinding, nonsignificant results obtained by this study may be related to insufficient power to detect differences because few trials reported appropriate blinding. Therefore, lack of reporting greatly hindered our ability to conclude, and thus, the influence of blinding on treatment effect estimates for this area is still unknown. Editors and reviewers can reinforce good reporting practice by following CONSORT guidelines and recommendations reported here to achieve good reporting of methodological indicators, so that accurate information can be extracted from trial reports. This is a crucial step to advance the field.

In addition, heterogeneity of the data set could be another factor that could explain in part the results of this study. According to Hempel,[12] when studies differ in risk of bias, it is likely that they also differ in other ways, both measured and unmeasured, which also might affect the ES found in the study. According to Devereaux et al.,[79] poor reporting of trials may have acted as a "marker" for other bias and may explain in part the heterogeneity seen in the meta-epidemiological studies investigating the association between biases and treatment effect estimates.

The research field of risk of bias assessment has been controversial. The meta-epidemiological studies' methodologies are in constant development, and thus, the failure to detect a statistically significant association between blinding and ES should not be

interpreted as meaning that an effect is not present. More empirical evidence is needed to determine which biases (methodological quality factors) are likely to influence reported ES and under which conditions. This information will be crucial for the field of knowledge synthesis, especially for researchers and consumers of systematic reviews when determining the validity of the existing evidence.

Strengths and Limitations

As far as our knowledge, this is the first study in PT that has exclusively and exhaustively evaluated individual components of blinding (e.g., participants, assessors, statisticians, and investigators), as well as to what they were blinded to (hypothesis, details of intervention, random assignment, analysis, etc.). In addition, this is the first meta-epidemiological study conducted in PT evaluating the influence of blinding on treatment effects.

Our study assessed only published reports and not actual trials. We did not contact the authors of the actual trials because contacting authors is time consuming, inconvenient, unpractical, and potentially costly, which limits the feasibility of this approach every time that a risk of bias assessment is conducted. We decided to analyze reports with the information provided by the authors, which is the most common way that evidence syntheses are performed and reflect a pragmatic approach to risk of bias assessments.

The lack of reporting of different types of blinding precluded conducting sensitivity analysis and also having enough power to find a difference when this could exist. We could not perform sensitivity analyses for blinding to study hypothesis, details of interventions, outcome measures, and outcome analysis. In addition, evaluation of the success of blinding was also not well reported, and thus, it was impossible to perform further analyses. Evaluation of the success of blinding in clinical trials has been extensively debated because it does not actually evaluate the success of blinding.[82] The implication of guessing allocation is not the same as that of knowing allocation.

CONCLUSIONS

Reporting and possibly conduct of blinding in PT trials were found to be poor. In addition, trials with inadequate blinding of assessor and participant and overall appropriateness of blinding tended to underestimate treatment effects, although the differences were not statistically significant. The lack of statistical significance between blinding and ES should not be interpreted as that an effect is not present. More empirical evidence is needed to determine which biases (methodological quality factors) are likely to influence reported ES and under which conditions. This information will be crucial for the field of knowledge synthesis, especially for researchers and consumers of systematic reviews when aiming to summarize the existing evidence appropriately. Researchers should look for creative solutions to avoid performance and detection bias when possible. In addition, researchers and journal editors should adhere to the CONSORT statement to minimize the poor reporting of biases to facilitate the research in this field. Clinicians implementing research into practice should pay attention to performance and detection biases due to lack of blinding when reading and interpreting the results of RCTs in PT area because these biases could alter the accuracy of the treatment effects.

ACKNOWLEDGMENTS

The authors thank the Alberta Research Center for Health Evidence at the University of Alberta and all research assistants (Annabritt Chisholm (AC) and Dion Pasichnyk (DP)) who helped with data collection.

REFERENCES

1. Delgado-Rodríguez M, Llorca J: Bias. *J Epidemiol Community Health* 2004;58:635–41
2. Gluud LL: Bias in clinical intervention research. *Am J Epidemiol* 2006;163:493–501
3. Schulz KF, Grimes DA: Blinding in randomised trials: hiding who got what. *Lancet* 2002;359:696–700
4. Kjaergard LL, Villumsen J, Gluud C: Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135:982–9
5. Nüesch E, Reichenbach S, Trelle S, et al: The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Care Res* 2009;61:1633–41
6. Wood L, Egger M, Gluud LL, et al: Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336:601–5
7. Schulz KF, Chalmers I, Hayes RJ, et al: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12
8. Savovic J, Jones HE, Altman DG, et al: Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38

9. Als-Nielsen B, Gluud L, Gluud C. Methodological quality and treatment effects in randomised trials: a review of six empirical studies. Paper presented at the 12th Cochrane Colloquium 2004; October 2–6, 2004; Ottawa, Ontario, Canada
10. Balk EM, Bonis PAL, Moskowitz H, et al: Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973–82
11. Bialy L, Vandermeer B, Lacaze-Masmonteil T, et al: A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. *Evid Based Child Health* 2014;9:1052–9
12. Hempel S Detection of associations between trial quality and effect sizes. 2012. Available at: http://www.ncbi.nlm.nih.gov/books/NBK84249/. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Jan. Report No.: 12-EHC010-EF. Accessed April 7, 2016
13. Hempel S, Suttorp M, Miles J. Empirical evidence of associations between trial quality and effect size. 2011. Available at: http://www.ncbi.nlm.nih.gov/books/NBK56925/. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011 Jun. Report No.: 11-EHC045-EF. Accessed April 7, 2016
14. Van Tulder MW, Suttorp M, Morton S, et al: Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine* 2009;34:1685–92
15. 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
16. Pildal J, Hróbjartsson A, Jørgensen KJ, et al: Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847–57
17. Hewitt CE, Kumaravel B, Dumville JC, et al: Assessing the impact of attrition in randomized controlled trials. *J Clin Epidemiol* 2010;63:1264–70
18. Kjaergard LL, Als-Nielsen B: Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *Br Med J* 2002;325:249
19. Trowman R, Dumville JC, Torgerson DJ, et al: The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *J Clin Epidemiol* 2007;60:1229–33
20. Boutron I, Tubach F, Giraudeau B, et al: Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 2004;57:543–50
21. Villamar MF, Contreras VS, Kuntz RE, et al: The reporting of blinding in physical medicine and rehabilitation randomized controlled trials: a systematic review. *J Rehabil Med* 2013;45:6–13
22. Fregni F, Imamura M, Chien HF, et al: Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: a report of the international placebo symposium working group. *Am J Phys Med Rehabil* 2010;89:160–72
23. Boutron I, Tubach F, Giraudeau B, et al: Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA* 2003;290:1062–70
24. Armijo-Olivo S, Fuentes C, Rogers T, et al: How should we evaluate the risk of bias of physical therapy trials?: A psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in Physical Therapy (PT) area: a study protocol. *Syst Rev* 2013;2:88
25. Armijo-Olivo S, Ospina M, da Costa BR, et al: Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9:e96920
26. World Confederation for Physical Therapy. Position statement: standards of physical therapy practice. World Confederation for Physical Therapy; 2011. Available at: http://www.wcpt.org/sites/wcpt.org/files/files/WCPT_Description_of_Physical_Therapy-Sep07-Rev_2.pdf. Accessed on April 7, 2016
27. Sterne JA, Jüni P, Schulz KF, et al: Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513–24
28. Herbison P, Hay-Smith J, Gillespie WJ: Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. *J Clin Epidemiol* 2011;64:1070–5
29. Siersma V, Als-Nielsen B, Chen W, et al: Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med* 2007;26:2745–58
30. Olivo SA, Macedo LG, Gadotti IC, et al: Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther* 2008;88:156–75
31. Friedberg JP, Lipsitz SR, Natarajan S: Challenges and recommendations for blinding in behavioral interventions illustrated using a case study of a behavioral intervention to lower blood pressure. *Patient Educ Couns* 2010;78:5–11
32. Armijo-Olivo S, Cummings G, Fuentes CJ, et al: Identifying items to assess methodological quality in physical therapy trials: a factor analysis. *Phys Ther* 2014;94:1272–84
33. Armijo-Olivo S, Saltaji H, da Costa BR, et al: What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. *BMJ Open* 2015;5:e008562
34. Nüesch E, Trelle S, Reichenbach S, et al: The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ* 2009;339:b3244

35. Nüesch E, Trelle S, Reichenbach S, et al: Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;341:c3515

36. Cohen J Chapter 1: the concepts of power analysis. In Cohen J (ed). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, Academic Press, Inc., 1988:1–17

37. Reichenbach S, Sterchi R, Scherer M, et al: Meta-analysis: chondroitin for osteoarthritis of the knee or hip. *Ann Intern Med* 2007;146:580–90

38. DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88

39. Spittle A, Orton J, Doyle LW, et al: Early developmental intervention programs post hospital discharge to prevent motor and cognitive impairments in preterm infants. *Cochrane Database Syst Rev* 2007

40. Fransen M, McConnell S, Hernandez MG, et al: Exercise for osteoarthritis of the hip. *Cochrane Database Syst Rev* 2009

41. Handoll-Helen HG, Cameron ID, Mak-Jenson CS, et al: Multidisciplinary rehabilitation for older people with hip fractures. *Cochrane Database Syst Rev* 2009

42. Harvey LA, Brosseau L, Herbert RD: Continuous passive motion following total knee arthroplasty in people with arthritis. *Cochrane Database Syst Rev* 2010

43. Katalinic OM, Harvey LA, Herbert RD, et al: Stretch for the treatment and prevention of contractures. *Cochrane Database Syst Rev* 2010;(9)

44. Manheimer E, Cheng K, Linde K, et al: Acupuncture for peripheral joint osteoarthritis. *Cochrane Database Syst Rev* 2010

45. Ostelo-Raymond WJG, Costa-Leonardo OP, Maher CG, et al: Rehabilitation after lumbar disc surgery. *Cochrane Database Syst Rev* 2008

46. Rutjes-Anne WS, Nüesch E, Sterchi R, et al: Therapeutic ultrasound for osteoarthritis of the knee or hip. *Cochrane Database Syst Rev* 2010

47. Rutjes-Anne WS, Nüesch E, Sterchi R, et al: Transcutaneous electrostimulation for osteoarthritis of the knee. *Cochrane Database Syst Rev* 2009

48. Schaafsma F, Schonstein E, Whelan KM, et al: Physical conditioning programs for improving work outcomes in workers with back pain. *Cochrane Database Syst Rev* 2010;(1)

49. Davies P, Taylor F, Beswick A, et al: Promoting patient uptake and adherence in cardiac rehabilitation. *Cochrane Database Syst Rev* 2010

50. Effing T, Monninkhof-Evelyn EM, Valk-Paul PDLP, et al: Self-management education for patients with chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2007

51. Puhan MA, Gimeno SE, Scharplatz M, et al: Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2009

52. Taylor RS, Dalal H, Jolly K, et al: Home-based versus centre-based cardiac rehabilitation. *Cochrane Database Syst Rev* 2010;(1)

53. Main E, Prasad A, van der Schans CP: Conventional chest physiotherapy compared to other airway clearance techniques for cystic fibrosis. *Cochrane Database Syst Rev* 2005;(1)

54. O'Brien K, Nixon S, Tynan AM, et al: Aerobic exercise interventions for adults living with HIV/AIDS. *Cochrane Database Syst Rev* 2010;(8)

55. Jolliffe J, Rees K, Taylor-Rod RS, et al: Exercise-based rehabilitation for coronary heart disease. *Cochrane Database Syst Rev* 2001;(1)

56. Shaw KA, Gennat HC, O'Rourke P, et al: Exercise for overweight or obesity. *Cochrane Database Syst Rev* 2006;(4)

57. Lacasse Y, Goldstein R, Lasserson TJ, et al: Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2010;(3)

58. Sirtori V, Corbetta D, Moja L, et al: Constraint-induced movement therapy for upper extremities in stroke patients. *Cochrane Database Syst Rev* 2009

59. States RA, Pappas E, Salem Y: Overground physical therapy gait training for chronic stroke patients with mobility deficits. *Cochrane Database Syst Rev* 2009;(3)

60. Mead GE, Morley W, Campbell P, et al: Exercise for depression. *Cochrane Database Syst Rev* 2009;(3)

61. Pollock A, Baer G, Pomeroy VM, et al: Physiotherapy treatment approaches for the recovery of postural control and lower limb function following stroke. *Cochrane Database Syst Rev* 2007;(1)

62. Mehrholz J, Friis R, Kugler J, et al: Treadmill training for patients with Parkinson's disease. *Cochrane Database Syst Rev* 2010;(1)

63. Woodford HJ, Price C: EMG biofeedback for the recovery of motor function after stroke. *Cochrane Database Syst Rev* 2007;(2)

64. Watson L, Ellis B, Leng GC: Exercise for intermittent claudication. *Cochrane Database Syst Rev* 2010;(1)

65. Kramer MS, McDonald SW: Aerobic exercise for women during pregnancy. *Cochrane Database Syst Rev* 2006;(3)

66. Orozco LJ, Buchleitner AM, Gimenez PG, et al: Exercise or exercise and diet for preventing type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2008;(3)

67. Bendermacher-Bianca LW, Willigendael EM, Teijink-Joep AW, et al: Supervised exercise therapy versus non-supervised exercise therapy for intermittent claudication. *Cochrane Database Syst Rev* 2006;(2)

68. Forster A, Lambley R, Hardy J, et al: Rehabilitation for older people in long-term care. *Cochrane Database Syst Rev* 2009;(1)

69. Cramp F, Daniel J: Exercise for the management of cancer-related fatigue in adults. *Cochrane Database Syst Rev* 2008;(2)

70. Moseley AM, Herbert RD, Maher CG, et al: Reported quality of randomized controlled trials of

physiotherapy interventions has improved over time. *J Clin Epidemiol* 2011;64:594–601

71. Boutron I, Guittet L, Estellat C, et al: Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007; 4:e61
72. Moroz A, Freed B, Tiedemann L, et al: Blinding measured: a systematic review of randomized controlled trials of acupuncture. *Evid Based Complement Alternat Med* 2013;2013:708251
73. Zhang CS, Tan HY, Zhang GS, et al: Placebo devices as effective control methods in acupuncture clinical trials: a systematic review. *PLoS One* 2015;10:e0140825
74. Fuentes J, Armijo-Olivo S, Funabashi M, et al: Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: an experimental controlled study. *Phys Ther* 2014;94:477–89
75. Hansson L, Hedner T, Dahlöf B: Prospective randomized open blinded end-point (PROBE) study. A novel design for intervention trials. Prospective Randomized Open Blinded End-Point. *Blood Press* 1992;1:113–9
76. Johnston BC, Da Costa BR, Devereaux PJ, et al: The use of expertise-based randomized controlled trials to assess spinal manipulation and acupuncture for low back pain: a systematic review. *Spine* 2008;33:914–8
77. Adamson J, Cockayne S, Puffer S, et al: Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemp Clin Trials* 2006;27:305–19
78. Silverman WA, Altman DG: Patients' preferences and randomised trials. *Lancet* 1996;347:171–4
79. Devereaux PJ, Choi PTL, El-Dika S, et al: An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol* 2004;57:1232–6
80. Moher D, Hopewell S, Schulz KF, et al: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012;10:28–55
81. Schulz KF, Altman DG, Moher D: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332
82. Sackett DL: Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;328:1136