

Identifying Items to Assess Methodological Quality in Physical Therapy Trials: A Factor Analysis

Susan Armijo-Olivo, Greta G. Cummings, Jorge Fuentes, Humam Saltaji, Christine Ha, Annabritt Chisholm, Dion Pasichnyk, Todd Rogers

S. Armijo-Olivo, BScPT, MScPT, PhD, CLEAR (Connecting Leadership and Research) Outcomes Research Program, Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, 3-48 Corbett Hall, Edmonton, Alberta, Canada T6G 2G4. Address all correspondence to Dr Armijo-Olivo at: sla4@ualberta.ca or susanarmijo@gmail.com.

G.G. Cummings, PhD, RN, FCAHS, CLEAR Outcomes Research Program, University of Alberta, Edmonton Clinic Health Academy, University of Alberta.

J. Fuentes, BPT, PhD, MSc, Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, and Department of Physical Therapy, Catholic University of Maule, Talca, Chile.

H. Saltaji, DDS, MSc(Ortho), Orthodontic Graduate Program, Faculty of Medicine and Dentistry, School of Dentistry, University of Alberta.

C. Ha, BSc, Rehabilitation Research Center, Faculty of Rehabilitation Medicine, University of Alberta.

A. Chisholm, BSc, Alberta Research Centre for Health Evidence, Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta.

Author information continues on next page.

Background. Numerous tools and individual items have been proposed to assess the methodological quality of randomized controlled trials (RCTs). The frequency of use of these items varies according to health area, which suggests a lack of agreement regarding their relevance to trial quality or risk of bias.

Objective. The objectives of this study were: (1) to identify the underlying component structure of items and (2) to determine relevant items to evaluate the quality and risk of bias of trials in physical therapy by using an exploratory factor analysis (EFA).

Design. A methodological research design was used, and an EFA was performed.

Methods. Randomized controlled trials used for this study were randomly selected from searches of the Cochrane Database of Systematic Reviews. Two reviewers used 45 items gathered from 7 different quality tools to assess the methodological quality of the RCTs. An exploratory factor analysis was conducted using the principal axis factoring (PAF) method followed by varimax rotation.

Results. Principal axis factoring identified 34 items loaded on 9 common factors: (1) selection bias; (2) performance and detection bias; (3) eligibility, intervention details, and description of outcome measures; (4) psychometric properties of the main outcome; (5) contamination and adherence to treatment; (6) attrition bias; (7) data analysis; (8) sample size; and (9) control and placebo adequacy.

Limitation. Because of the exploratory nature of the results, a confirmatory factor analysis is needed to validate this model.

Conclusions. To the authors' knowledge, this is the first factor analysis to explore the underlying component items used to evaluate the methodological quality or risk of bias of RCTs in physical therapy. The items and factors represent a starting point for evaluating the methodological quality and risk of bias in physical therapy trials. Empirical evidence of the association among these items with treatment effects and a confirmatory factor analysis of these results are needed to validate these items.



Post a Rapid Response to this article at: ptjournal.apta.org

D. Pasichnyk, MSc(Epidemiology), Quality Management in Clinical Research, University of Alberta.

T. Rogers, PhD, Centre for Research in Applied Measurement and Evaluation, University of Alberta.

[Armijo-Olivo S, Cummings GG, Fuentes J, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. *Phys Ther.* 2014;94:1272–1284.]

© 2014 American Physical Therapy Association

Published Ahead of Print: May 1, 2014

Accepted: April 14, 2014

Submitted: October 8, 2013

Methodological quality assessment of health research has been a matter of interest and research since the concept of evidence-based practice (EBP) was introduced in 1992 and the generation of systematic reviews and knowledge synthesis research began. Methodological quality assessment is of paramount importance for health researchers and policy and decision makers because only the best quality evidence is recommended for uptake and to guide recommendations for future research and clinical practice.^{1,2}

Although quality assessment has been acknowledged to be an important part of knowledge synthesis and has evolved in many aspects since its inception (ie, changes in definition, methods, tools), it has been a controversial topic for many years. Recently, the Cochrane Collaboration has proposed a shift in the approach to quality assessment. The concept of trial quality has been linked to the internal validity of the study or risk of bias. Thus, a study has better quality when it has a lower risk of bias.³

Inconsistencies in the approaches to quality assessment have been discussed by many researchers.^{1,2,4–6} Numerous tools and items contained in these tools have been found in all areas of health research.^{2,6} Specific to rehabilitation and physical therapy trials, 7 tools have been identified: Delphi list, Physiotherapy Evidence Database (PEDro), Maastricht scale, Maastricht-Amsterdam list, Bizzini scale, van Tulder scale, and Jadad scale.² These tools have not been adequately developed and have not been adequately tested for validity and reliability in physical therapy research. Furthermore, the link between many items contained in these tools and bias is unclear because some items may relate more to the adequacy of reporting than

to methodological quality.^{7,8} In our recent study,⁷ the frequency of use of these items (ie, how many times these items are used in different tools to evaluate quality of randomized controlled trials [RCTs]) varied according to health field (ie, general health research and physical therapy research). This finding suggests a lack of agreement regarding item relevance to trial quality or risk of bias. Our results called for an in-depth analysis of the items used to determine trial quality and risk of bias of RCTs to provide evidence of validity for these items.

Evidence of validity consists of (or is the sum of) many different types of evidence, such as content, criterion or criteria, construct, and other related evidence.⁹ One method to provide evidence of validity of the items is to evaluate whether or to what extent the items are associated with treatment effects. In the area of knowledge synthesis, this evaluation has been done using meta-epidemiological approaches (ie, “empirical evidence”). For example, The Cochrane Collaboration has used empirical evidence to justify the components and items contained in its Risk of Bias Tool.^{3,10} The Risk of Bias Tool includes 6 domains: sequence generation, allocation concealment, blinding, missing outcome data, selective outcome reporting, and “other sources of bias” (eg, early stopping for benefit, design-specific features such as adequate wash-out period in crossover trials). For example, inadequate allocation concealment and lack of double blinding can lead to overestimation of treatment effects by an average of 18% and 9%, respectively.^{11–13} Other factors, such as the method of randomization,^{14,15} follow-up proportions,^{16,17} and industry sponsorship,^{18,19} also have influenced the results of trials. All of these factors can lead to overestimates of treatment effects, or bias, at the trial level and thereby to biased

or inaccurate results and conclusions in systematic reviews and meta-analyses.^{13,15,16,20,21}

Another way to provide evidence of validity is through psychometric evaluation of the tools and their items (ie, content, criterion, and construct validity). Factor analysis is one of the most frequently used methods to determine relevance of items. Factor analysis has been used to determine items of tools to evaluate the quality of RCTs for some specific health areas such as dermatology²² or general health research²³ and to validate items used to evaluate the quality of systematic reviews.²⁴ However, this method has not been conducted in other health areas, specifically in the area of physical therapy.⁷ To our knowledge, there is no evidence of the underlying component structure of items and relevant items from tools used to evaluate the methodological quality and risk of bias of physical therapy trials. Therefore, our main objective was to identify the latent structure of these 45 items from the 7 existing tools used to evaluate the methodological quality of RCTs in the physical therapy field. Based on this main objective, exploratory factor analysis (EFA) was the best choice. Based on the literature,^{25,26} EFA aims to: (1) identify the factor structure or model for a set of variables (ie, the number of factors and pattern of the factor loadings), (2) determine whether the factors are correlated, and (3) name factors obtained.²⁶ All of these objectives were of interest to us.

Physical therapy interventions are classified as complex interventions²⁷ and have diverse methodological and clinical aspects that may affect trial results, such as the type and intensity of therapy, type of approach (ie, standardized or individually tailored), and the skills and experience of therapists. In addition, because of the nature of physical therapy inter-

ventions (eg, manual therapy, exercises), blinding of the therapists and patients is not always possible. Appropriate blinding of participants and all key study personnel, therefore, is unlikely to be accomplished for most physical therapy and other nonpharmacological trials. Blinding of outcome assessment, however, has been used as a proxy quality measure without validation. Thus, assessment of physical therapy trials may need to consider not only general components of design (eg, randomization, concealment, blinding) but also more specific components, such as type and intensity of therapy, type of intervention approach (ie, standardized or tailored), and the skills and experience of therapists. However, whether these factors are valid to measure the quality and risk of bias of an RCT in the physical therapy field is not yet known.

In order to guide quality or risk of bias assessments to appropriately inform decision making, it is important to know which items should be included in these tools based on psychometric evaluation as well as empirical evidence. These 2 types of analyses are complementary. This information is urgently needed to develop guidelines for the design, conduct, report, and implementation of trials. In addition, this information is important for systematic reviewers and meta-analysts to evaluate the quality of intervention trials.

Method Studies Included

Randomized controlled trials were obtained by searching the Cochrane Database of Systematic Reviews, using the key words “physical therapy” or “physiotherapy,” “rehabilitation,” “exercise,” “electrophysical agents,” “acupuncture,” “massage,” “transcutaneous electrical nerve stimulation,” “interferential current,” “ultrasound,” “stretching,” “chest therapy,” “pulmonary rehabilitation,”

“manipulative therapy,” “mobilization,” and related terms. The Cochrane Database of Systematic Reviews was used because systematic reviews conducted by the Cochrane Collaboration in the physical therapy field have been recognized as scientifically and more rigorously conducted than non-Cochrane reviews.²⁸ In addition, the Cochrane Database of Systematic Reviews provides a high level of detail and consistency of reporting across meta-analyses. Meta-analyses and their trials were included if: (1) they included at least 5 RCTs comparing at least 2 interventions, at least 1 of which is currently or potentially part of physical therapist practice according to the World Confederation for Physical Therapy,²⁹ and (2) the allocation of participants to interventions in the RCTs was random or reported to be random.

A unique code generated by the Reference Manager bibliographic program (The Thomson Corporation, Philadelphia, Pennsylvania) was assigned to each meta-analysis and trial that met the inclusion criteria. This code was used to randomly select studies to be analyzed for this analysis and to randomize the order of evaluation. The first author (S.A.O.) randomly selected each meta-analysis to be included and accompanying trials by drawing the code of the selected meta-analysis first and then from each trial from an opaque envelope. This process ensured that the researcher had no influence on the studies selected and no influence on the order of evaluation. According to Stevens,²⁶ 200 trials would be sufficient to perform an EFA.

Identification of Items

A total of 214 randomly selected RCTs were evaluated using the 45 items, selected from 7 tools (ie, Delphi list, PEDro scale, Maastricht

scale, Maastricht-Amsterdam list, Bizzini scale, van Tulder scale, and Jadad scale) that are most commonly used or reported to be valid in the physical therapy field. Items were selected such that all unique items from these tools were included.² Details of items used for analyzing the quality and risk of bias of selected RCTs are provided in Table 1.

The definitions of items (ie, how they are defined and assessed) were obtained from the guidelines of the original tools. A 3-category response (“yes,” “no,” “unclear”), which was the most common response format in the original scales, was used. Thus, standardized guidelines for assessing the items were compiled and distributed to all reviewers before starting training and data collection process.

Reviewers

A review panel consisting of 6 reviewers with experience in different areas of health sciences research participated in this study. Two reviewers had bachelor's degrees in health sciences, 1 had a master's degree in public health, 1 had a master's degree in dentistry and was currently working on a PhD in orthodontics, and 2 were physical therapists and had master's degrees and PhDs in rehabilitation sciences.

Reviewer Training

All reviewers received the same training and standardized guidelines for assessing the studies (as previously mentioned). Reviewer training was carried out with 10 studies not included in the set of studies to be reviewed. Each of the 10 training studies was independently reviewed by each team member and discussed by all reviewers in a group meeting to determine consistency in ratings. The first author performed the training for all reviewers. The training lasted approximately 1 month. In addition, the team members met on

a regular basis to discuss ratings of studies performed by all reviewers. These studies were not included in the analyses. These meetings also were performed to increase consistency in ratings and to determine if there were any issues regarding the process of data extraction and quality assessment.

Data Extraction and Quality Assessment Process

During the data extraction phase, each study was independently evaluated by 2 members of the panel following standardized guidelines distributed to each reviewer. We developed and pilot tested an electronic form for data extraction. Data on methodological quality for each RCT were extracted and entered directly into the electronic form using Microsoft Access (Microsoft Corp, Redmond, Washington).

The 2 reviewers who assessed the same study compared their assessments. Any discrepancies were resolved by discussion between the 2 reviewers. If a consensus rating was not achieved, the 2 reviewers consulted with a third reviewer (first author). The full consensus rating between the 2 reviewers analyzing the same study was used for all analyses.

Data Analysis

Exploratory factor analysis was used to identify the latent structure of the 45 items from the 7 existing tools used to evaluate the methodological quality of RCTs in the physical therapy field. First, the 45 items were examined for variability across the 3 response options. Second, the Kaiser-Guttman rule (number of components with eigenvalues ≥ 1 yielded by a principal components extraction), the scree test, and Kaiser's image factoring followed by varimax rotation were used to identify the number of common factors that underlie the structure of items

after step 1.^{25,26} Third, following identification of the number of common factors, the items were subjected to an EFA using principal axis extraction followed by a varimax rotation and an oblique transformation. Correlations among factors were analyzed to determine whether a varimax or oblique transformation would be used. Items that did not load on any of the retained factors or with factor loading ≤ 0.36 were then sequentially removed based on the recommendation provided by Stevens.^{25,26} He suggested that a loading of 0.722 can be considered significant for a sample size of 50, a loading greater than 0.512 is significant for a sample size of 100, and a loading greater than 0.364 is significant for a sample size of 200. The solution that best represented simple structure and that was interpretable was selected. SPSS version 17 software (SPSS Inc, Chicago, Illinois) was used to perform all analyses. After conducting the EFA, the retained factors were named by the first author and then discussed and verified by the members of the review panel.

Interpretation of Factor Solution

To determine the names of the factors that underlie the structure of items used in physical therapy tools and the interpretability of the factors obtained from the EFA, the first author made the initial classification based on a paradigm shift of methodological quality to risk of bias introduced by The Cochrane Collaboration.^{3,10} Thus, the naming of factors when possible was linked to risk of bias according to standard classification and guidelines.^{30,31} After this naming, the review panel provided feedback by examining the coherence of the names given to the grouping factors. Reviewers considered each factor by asking, “What type of threats to validity (bias) or precision is each particular factor addressing?”^{10,30} or “What are the

Assessing Methodological Quality in Physical Therapy Trials

Table 1.

Items to Measure the Methodological Quality of Randomized Controlled Trial in Physical Therapy Area Considered in Factor Analysis

	Items Included in the Scales	Included in Factor Analysis Solution	Reporting	Conduction
	Patient Selection (Inclusion and Exclusion and Description of Participants)			
1	Inclusion criteria clearly defined/eligibility criteria specified	✓	X	
2	Exclusion criteria clearly defined/eligibility criteria specified	✓	X	
3	Baseline comparability (group equivalence, homogeneity) regarding the most important prognostic indicators	✓		X
	Assignment, Randomization, and Allocation Concealment			
4	Study described as randomized	✓	X	
5	Randomization method performed	✓		X
6	Method of randomization described and appropriate	✓		X
7	Method of randomization concealed	✓		X
	Blinding			
8	Study described as double blind	✓	X	
9	Method of blinding appropriate	✓		X
10	Blinding of investigator	✓	X	X
11	Blinding of assessor	✓		X
12	Observer blinding evaluated and successful ^a			X
13	Blinding of participants/patients	✓	X	X
14	Patient blinding evaluated and successful ^a			X
15	Blinding of therapists/care provider	✓	X	X
16	Therapist blinding evaluated and successful ^a			X
17	Blinding of the outcome (data analyst)	✓	X	X
	Interventions			
18	Treatment protocol adequately described for the treatment group regarding type of intervention; duration of each intervention; frequency, intensity, and dosage	✓	X	
19	Treatment protocol adequately described for the control group regarding type of intervention; duration of each intervention; frequency, intensity, and dosage	✓	X	
20	Treatment protocol adequately described for the comparison group regarding type of intervention; duration of each intervention; frequency, intensity, and dosage (if applicable) ^a		X	
21	Control adequate (presence of a true control group)	✓		X
22	Placebo adequate (presence of a true placebo group)	✓		X
23	Cointerventions avoided or comparable	✓		X
24	Cointerventions reported for each group separately	✓	X	
25	Testing of participant adherence to treatment protocol (report of adherence) ^b		X	X
26	Adherence acceptable in all groups (80% of treatment received)	✓		X

(Continued)

Table 1.
Continued

	Items Included in the Scales	Included in Factor Analysis Solution	Reporting	Conduction
	Attrition, Follow-up, and Protocol Deviation			
27	Report of withdraws and dropouts (rate)	✓	X	
28	Withdrawal/dropout rate acceptable (<20%)	✓		X
29	Reasons for withdrawals and dropouts reported	✓	X	
30	Adverse effects described	✓	X	
31	Short follow-up measurement performed ^b			X
32	Long-term follow-up measurement performed ^a			X
	Outcomes			
33	Outcome measures described ^b		X	
34	Relevant outcomes were included (there was a good rationale between intervention and outcome) ^b			X
35	Validity for main outcome measures reported	✓	X	X
36	Responsiveness for main outcome measures reported	✓	X	X
37	Reliability for main outcome measures reported	✓	X	X
38	Use of objective outcome measures (use of measurements that can be scored) ^b			X
39	The timing of the outcome assessment was comparable in all groups (same time)	✓		X
	Statistical Analysis			
40	Descriptive measures (point estimates and measures of variability) identified and reported for the primary outcome	✓	X	
41	Appropriate statistical analysis used	✓		X
42	Sample size calculation performed prior to initiation of the study	✓		X
43	Adequate sample size	✓		X
44	Sample size described for each group	✓	X	
45	Intention-to-treat analysis used ^b			X

^a Items deleted because of low variability before performing factor analysis.

^b Items not loaded in any of the factors.

grouping factors intending to capture?” Thus, factors were classified into the threats to validity or precision that best represented the concepts being addressed. Disagreements in classification of factors were resolved by consensus. Full consensus was used to name the factors.

Results

The Figure shows the process of identifying studies to be included in the factor analysis. A total of 214 randomly selected RCTs were evaluated using the 45 items.

Before performing the factor analysis, 5 items were excluded because they had no variability or they were not applicable for many studies. The 5 items were: observer blinding evaluated and successful, participant blinding evaluated and successful, therapist blinding evaluated and successful, long-term follow-up measurement performed, and description of the intervention for a third comparison group. Thus, 40 items were included in the analysis. Table 1 shows the items included in the factor analysis and those that were excluded based on lack of variability

or because they did not load in the final solution.

After applying the principal component extraction, the number of factors suggested by the Kaiser-Guttman rule was 16; the scree test and image plus varimax suggested the number was 8. A principal axis extraction followed by a varimax rotation and an oblique transformation were then completed. Correlations among pairs of oblique factors were all low ($\leq .10$); therefore, varimax solutions were retained.

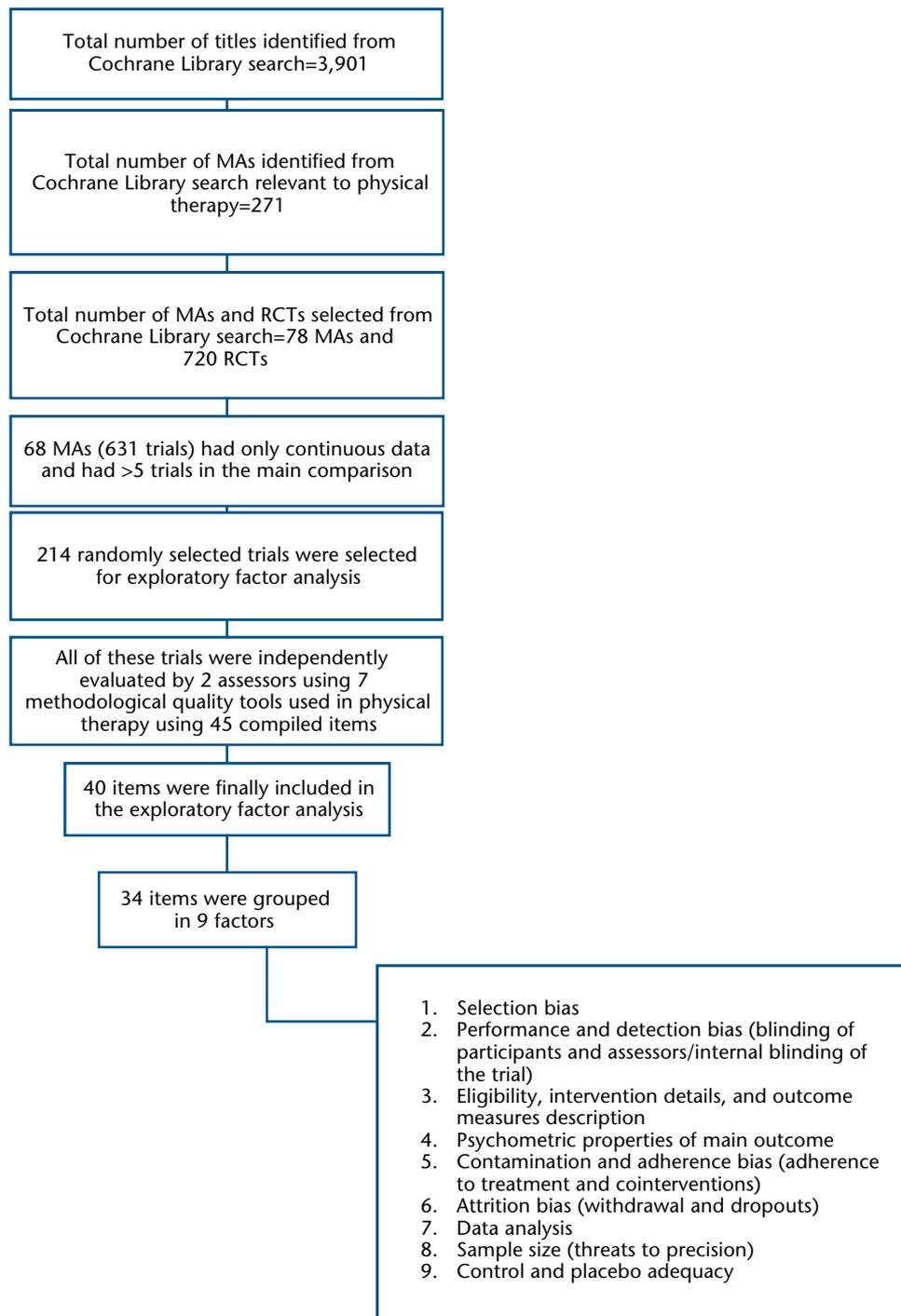


Figure. Diagram for identification of studies. MA=meta-analysis, RCT=randomized controlled trial.

Different factor solutions were evaluated regarding single structure and interpretability of the factors. This process included several iterations of the analysis. Inspection of the final

solution with 8 factors revealed that several items did not load on any of the 8 factors and that others had very low loadings. These items were sequentially removed to the point

where 34 items remained. Seven-, 8-, and 9-factor solutions were then obtained to determine which of these solutions had the simplest structure and was clearly interpreta-

ble as described above. Although the 8- and 9-factor solutions possessed nearly equal simple structure, the 9-factor solution was more interpretable than the 8-factors solution, based on our knowledge and theoretical grounds. A copy of the 9-factor solution and items loaded in each factor is provided in Table 2. As shown in the table, the 9 factors named by the research team were: (1) selection bias (5 items); (2) performance and detection bias (blinding of participants and assessors, internal blinding of the trial) (4 items); (3) eligibility, interventions details, and description of the outcomes measures (5 items); (4) psychometric properties of main outcome (4 items); (5) contamination and adherence bias (adherence to treatment and cointerventions) (3 items); (6) attrition bias (withdrawal and dropouts) (4 items); (7) data analysis (4 items); (8) sample size (threats to precision) (2 items); and (9) control and placebo adequacy (3 items).

Discussion

The main results of this study show that 45 items from tools used in the physical therapy field² could be reduced to 34 items that loaded on 9 independent common factors, which possessed a simple structure and were interpretable. To our knowledge, this type of validity evidence has not been conducted in the physical therapy field.⁷ Therefore, this study provides novel information regarding the factors that underlie the structure of items included in tools to evaluate trial quality commonly used in the physical therapy field.

Our results will be valuable to a number of stakeholders, including researchers, systematic reviewers and meta-analysts, methodologists, clinicians, and policy makers working in the field of physical therapy. These results are a starting point to

determine items that are important to determine trial quality in the area of physical therapy when analyzing individual physical therapy trials, performing a systematic review, or searching for high-quality evidence for decision making. This discussion concentrates on the main findings of the factor analysis solution, highlighting the organization of factors and relevance of factors to the physical therapy field, discrepancies between the factor analysis solution and team thoughts, and limitations of this study.

Organization of Factor Solution and Relevance to the Physical Therapy Field

According to the factors obtained through factor analysis, it is possible to recognize that physical therapy items are grouped very closely to those proposed in the Cochrane Risk of Bias Tool to evaluate risk of bias of RCTs in health research. As mentioned previously, the Risk of Bias Tool includes 6 domains: sequence generation, allocation concealment, blinding, missing outcome data, selective outcome reporting, and “other sources of bias.” Sequence generation, allocation concealment, blinding, and missing outcome data also are factors describing the items from RCTs in the physical therapy field. However, tools used to evaluate RCTs in the physical therapy field also included items related to description of treatment (ie, treatment fidelity) and items linked to adherence and contamination bias with treatment, which were shown to be important based on our factor analysis results. Physical therapy interventions are classified as complex interventions^{2,27} comprising diverse aspects that may affect trial results, such as type of therapy and its intensity, type of approach (standardized or individually tailored), and the skills and experience of therapists. Thus, based on the factor analysis, these items should be con-

sidered when evaluating the methodological quality or risk of bias of RCTs in the physical therapy field.

Other methodological components within the Risk of Bias Tool and physical therapy tools that have traditionally been used to determine trial quality in health research have not been investigated empirically (ie, they have not been investigated using a meta-epidemiological approach); thus, the evidence base is restricted and incomplete. Therefore, we recommend that research evidence be expanded to different health areas regarding the association among methodological factors (items used in quality tools) and their link to treatment estimates, especially those that involve complex interventions such as allied health areas and physical therapy.

The organization of the items on the 9 factors found in the present study aligns closely to what we had proposed in a previous study describing the frequency of items and their categorizations.⁷ In our previous study,⁷ we organized the items used in physical therapy tools into 7 groups: (1) patient selection (inclusion and exclusion criteria, description of study participants); (2) assignment, randomization, and allocation concealment; (3) blinding; (4) interventions; (5) attrition, follow-up, and protocol deviations; (6) outcomes; and (7) statistical analysis. The 2 additional factors identified in our factor analysis were control and placebo adequacy and contamination and adherence (adherence to treatment and cointerventions). We had previously considered both of these factors under the category “interventions.”⁷ Thus, the factor analysis subdivided the intervention category into 3 different factors: (1) interventions details, (2) contamination and adherence to treatment bias, and (3) control and placebo adequacy. These 3 domains may require more

Table 2. Rotated Component Matrix Displaying the 9 Factors and Loadings for Items Used to Evaluate the Methodological Quality and Risk of Bias of Physical Therapy Trials^a

Item Description	Components								
	Selection Bias	Performance and Detection Bias (Blinding of Participants and Assessors/Internal Blinding of the Trial)	Eligibility, Intervention Details, and Outcome Measures Description	Psychometric Properties of Main Outcome	Contamination and Adherence Bias (Adherence to Treatment and Cointerventions)	Attrition Bias (Withdrawal and Dropouts)	Data Analysis	Sample Size (Threats to Precision)	Control and Placebo Adequacy
Blinding statistician	.783								
Blinding investigator	.749								
Described as randomized	.721								
Randomization concealed	.592								
Method of randomization described and appropriate	.472								
Blinding of assessors		.804							
Method of blinding appropriate		.759							
Blinding of participants		.751							
Study described as double blind		.392							
Exclusion criteria clearly defined			.726						
Treatment protocol adequately described for the treatment group regarding type of intervention; duration of each intervention; frequency, intensity, and dosage			.710						
Inclusion criteria clearly defined			.688						
Outcome measures described			.460						
Treatment protocol adequately described for the control group regarding type of intervention; duration of each intervention; frequency, intensity, and dosage			.430						
Validity for main outcome measures reported				.778					
Reliability for main outcome measures reported				.750					
Responsiveness for main outcome measures reported consensus				.489					

(Continued)

Table 2.
Continued

Components									
Item Description	Selection Bias	Performance and Detection Bias (Blinding of Participants and Assessors/Internal Blinding of the Trial)	Eligibility, Intervention Details, and Outcome Measures Description	Psychometric Properties of Main Outcome	Contamination and Adherence Bias (Adherence to Treatment and Cointerventions)	Attrition Bias (Withdrawal and Dropouts)	Data Analysis	Sample Size (Threats to Precision)	Control and Placebo Adequacy
The timing of the outcome assessment was comparable in all groups (same time)				-.412					
Cointerventions reported for each group separately					.828				
Cointerventions avoided/comparable					.824				
Adherence acceptable in all groups (80% of treatment received)					.404				
Report of withdrawals and dropouts (rate)						.766			
Reasons for withdrawals and dropouts reported						.745			
Sample size described for each group						.464			
Withdrawal/dropouts rate acceptable (<20%)						.408			
Baseline comparability (group equivalence, homogeneity) regarding the most important prognostic indicators							.686		
Adverse effects described							.609		
Appropriate statistical analysis used							.540		
Were point estimates and measures of variability identified and reported for the primary outcome?							.446		
Adequate sample size								.744	
Sample size calculations								.698	
Blinding of therapist									.638
Placebo adequate (presence of a placebo group)									.619
Control adequate (presence of a true control group)									-.379

^a Rotation method: varimax with Kaiser normalization. Rotation converged in 8 iterations.

attention when evaluating the methodological quality and risk of bias of physical therapy trials. Because physical therapy trials are much more complex than a pharmacological RCT, physical therapy-related tools used to measure methodological quality and risk of bias of primary RCTs in the physical therapy field should take into account not only adherence and standardization of the treatment protocol but also the precise performance of the intervention (treatment fidelity).² The next important step is to assess whether these 3 factors identified in the factor analysis are associated with treatment effect estimates.

Factor Solution and Discrepancies With Team Thoughts

According to the research team, of all of the items not included in the factor analysis solution, 1 item (intention-to-treat [ITT] analysis performed) could be considered important in evaluating the quality of RCTs in the physical therapy field based on theoretical grounds regarding methodological quality and risk of bias of RCTs in other health areas.^{17,32-34} Effect sizes from trials that excluded participants in their analysis or considered a modified ITT tended to be more beneficial than those from trials without exclusions, demonstrating that the ITT principle is important in preserving the benefits of randomization and keeping unbiased estimates when the objective of the trial is to determine treatment effectiveness.^{17,33,35} That is, biased results may be obtained if the comparability between the groups is lost when ITT is not used. However, some researchers argue that the choice of which approach to use to conduct or analyze clinical trials (effectiveness versus efficacy approach) depends on the objectives of the trial and on who is expected to utilize the results.³⁵ Thus, an ITT is not necessarily the analysis of choice in all

trials. The need to do an ITT analysis (effectiveness approach) or a per-protocol (PP) or as-treated (AT) analysis (efficacy approach) is based on the question that researchers want to answer. When investigators want to know the effect of a certain treatment under ideal conditions on patients who are adherent to treatment, an AT analysis or PP analysis (efficacy analysis) should be used. However, when researchers want to know whether the treatment works in clinical and practical conditions, an ITT analysis (effectiveness analysis) should be conducted. Other researchers have suggested that a “sensitivity analysis” (ie, analyzing data through 2 or more different methods [eg, using ITT analysis and PP or AT analysis]) should be conducted in order to test the validity of the conclusions.³⁵ However, it is unknown if this item can be linked to affect estimates in physical therapy trials. Research investigating the influence of the ITT principle on treatment estimates in physical therapy trials is warranted.

Two items within 2 factors loaded negatively, contrary to our expectations. These negative loadings were due to the scores that the analyzed studies received for these items. For example, the item “timing of the outcome assessment was comparable in all groups” was scored mainly as “yes” by 209 of the articles (98%). In contrast, the items dealing with validity, reliability, and responsiveness within the same factor were scored mainly as “no.” Thus, 74%, 68%, and 97% of the analyzed trials scored validity, reliability, and responsiveness as not accomplished, respectively. Therefore, these items behaved similarly for most of the trials analyzed (scored “no”) and “timing of the outcome assessment” behaved in the opposite direction (scored mainly “yes”). Therefore, it came with a negative loading. Thus, these loadings are an expression of

the way the analyzed trials behaved when these items were scored.

Based on the factor analysis, some items related more to “reporting quality” than to “conduct.” We defined methodological quality as “the confidence that the trial design, conduct, and analysis has minimized or avoided biases in its treatment comparisons”^{5(p63)} (eg, allocation concealment was appropriate). We defined quality of reporting as authors providing “information about the design, conduct, and analysis of the trial”^{5(p63)} (eg, method for concealing allocation was reported) (14 items, Tab. 1). Many of the tools used to evaluate methodological quality and risk of bias of health research have several items linked to reporting instead of conduct.⁷ This finding has been highlighted in previous research from our team and others.^{7,8} It is possible that reporting quality could be used as a proxy for trial quality; however, although quality of reporting is necessary to assess quality of conduct, quality of reporting can hide differences in quality of conduct and can actually underestimate or overestimate trial quality.^{36,37} Therefore, empirical evidence investigating the association between these items and treatment effect estimates still is needed in order to provide more validity evidence for the use of these items when evaluating the methodological quality and risk of bias of physical therapy trials.

Naming of Factors

We acknowledge that naming and classifying factors is a complex task and can be subjective, especially when items load on factors that were not anticipated. For example the item, “sample size described for each group” loaded on “attrition bias factor” instead of “sample size.” It may be that this item loaded on “attrition bias” because, in the process of determining how many participants

drop out or complete the trial, it is necessary to report the number of participants per group. On the other hand, the factor representing “sample size” included items measuring sample size calculation and its adequacy.

We performed this task (ie, naming factors) in duplicate and based on our ability, experience, and knowledge; however, the precise classification of factors could be debatable until empirical evidence supports the link of the items with specific bias. Moreover, according to MacCallum, “Models at their best can be expected to provide only a close approximation to observed data, rather than exact fit. . . . One can conclude only that the particular model is a plausible one.”^{38(p17)} Thus, we feel that this factor solution could be a working set of items and a starting point to determine items that could be used to evaluate the quality and risk of bias of RCTs in the physical therapy field. Nevertheless, this set of items and factors needs further validation using data from another set of trials. Thus, future research could explore performing a confirmatory factor analysis of these results.

Limitations

Although this study used a factor analysis based on rigorous methods and a robust number of trials ($n=214$), some limitations should be acknowledged. First, it may be possible that the trials selected for analysis were not representative of all physical therapy trials. However, a random sampling was used to decrease selection bias. Second, because of the exploratory nature of the results, the factor solution obtained can be applicable only to this set of trials. A confirmatory factor analysis using another set of trials needs to be performed in order to validate this model.

To our knowledge, this is the first factor analysis to explore the underlying component items used to evaluate methodological quality and risk of bias of physical therapy trials based on a robust number of trials ($n=214$). Therefore, this study provides novel evidence regarding the number of factors that underlie the structure of items included in tools frequently used in the physical therapy field to assess the methodological quality of RCTs in the physical therapy field. The items and factors represent a starting point for evaluating the methodological quality and risk of bias in physical therapy trials. Empirical evidence of the association between these items and treatment effects is needed to validate these items before widespread use. In addition, future research could explore performing a confirmatory factor analysis of these results.

Dr Armijo-Olivo, Dr Cummings, and Dr Rogers provided concept/idea/research design. Dr Armijo-Olivo, Dr Rogers, Dr Fuentes, and Dr Cummings provided writing. Dr Armijo-Olivo, Dr Fuentes, Dr Saltaji, Ms Ha, Ms Chisholm, and Mr Pasichnyk provided data collection. Dr Armijo-Olivo, Dr Cummings, and Dr Rogers provided data analysis. All authors provided data interpretation. Dr Armijo-Olivo provided project management and fund procurement. Dr Cummings, Dr Fuentes, Ms Chisholm, Ms Ha, and Mr Pasichnyk provided consultation (including review of manuscript before submission). The authors thank the Alberta Research Center for Health Evidence at the University of Alberta and all research assistants who helped with data collection.

This manuscript was presented at the 20th Cochrane Colloquium; September 19–23, 2013; Quebec, Canada, and at the Knowledge Translation Summer Institute; June 17–19, 2013; Hamilton, Ontario, Canada.

Dr Armijo-Olivo is supported by the Canadian Institutes of Health Research (CIHR) as a Banting Postdoctoral Fellow (Ottawa, Ontario, Canada), the Alberta Innovates Health Solutions (AIHS, Edmonton, Alberta, Canada), the STIHR Training Program from Knowledge Translation Canada, and the University of Alberta. Dr Cummings is funded both provincially and nationally and holds a

Population Health Investigator Award from the Alberta Heritage Foundation for Medical Research (2006–2013). She holds a Centennial Professorship at the University of Alberta (2013–2020). Dr Fuentes is supported by the Government of Chile, University of Alberta, through a dissertation fellowship and the Catholic University of Maule. Dr Saltaji is supported through a Clinician Fellowship Award by the AIHS, the Honorary Izaak Walton Killam Memorial Scholarship by the University of Alberta, and the WCHRI Award by the Women and Children’s Health Research Institute (WCHRI).

This project is funded by the Physiotherapy Foundation of Canada through a B.E. Schnurr Memorial Fund Award, by the AIHS through a knowledge translation initiative grant, by the Knowledge Translation Canada Research Stipend Program, by the CIHR Banting Program, and by the University of Alberta.

DOI: 10.2522/ptj.20130464

References

- 1 Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995; 16:62–73.
- 2 Olivo SA, Macedo LG, Gadotti IC, et al. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther*. 2008;88:156–175.
- 3 Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928
- 4 Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. *J Clin Epidemiol*. 2006;59:1249–1256.
- 5 Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054–1060.
- 6 Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol*. 2004;4:22. Available at: <http://www.biomedcentral.com/1471-2288/4/22>.
- 7 Armijo-Olivo S, Fuentes J, Ospina M, et al. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. *BMC Med Res Methodol*. 2013;13:116.
- 8 Dechartres A, Charles P, Hopewell S, et al. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol*. 2011;64:136–144.

- 9 Streiner D, Norman G. Validity. In: Streiner D, Norman G, eds. *Health Measurements Scales*. Oxford, United Kingdom: Oxford University Press; 2004:172-193.
- 10 Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available at: <http://www.cochrane-handbook.org>.
- 11 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; 352:609-613.
- 12 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408-412.
- 13 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: 601-605.
- 14 Berger VW, Weinstein S. Ensuring the comparability of comparison groups: is randomization enough? *Control Clin Trials*. 2004;25:515-524.
- 15 Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *J Clin Epidemiol*. 2007;60:1229-1233.
- 16 Hewitt CE, Kumaravel B, Dumville JC, Torgerson DJ. Assessing the impact of attrition in randomized controlled trials. *J Clin Epidemiol*. 2010;63:1264-1270.
- 17 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;339:679-683.
- 18 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA*. 2003;289:454-465.
- 19 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003;326:1167-1170.
- 20 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; 36:847-857.
- 21 Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *BMJ*. 2002;325:249.
- 22 Revuz J, Moysé D, Poli F, et al. A tool to evaluate rapidly the quality of clinical trials on topical acne treatment. *J Eur Acad Dermatol Venereol*. 2008;22:800-806.
- 23 Crowe M, Sheppard L. A general critical appraisal tool: an evaluation of construct validity. *Int J Nurs Stud*. 2011;48:1505-1516.
- 24 Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
- 25 Field AP. Exploratory factor analysis. In: Field AP, ed. *Discovering Statistics Using SPSS: (and Sex and Drugs and Rock 'n' Roll)*. Thousand Oaks, CA: Sage Publications; 2009:619-680.
- 26 Stevens J. Exploratory and confirmatory factor analysis. In: Stevens J, ed. *Applied Multivariate Statistics for the Social Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:385-470.
- 27 Kunz R, Autti-Rämö I, Anttila H, et al. A systematic review finds that methodological quality is better than its reputation but can be improved in physiotherapy trials in childhood cerebral palsy. *J Clin Epidemiol*. 2006;59:1239-1248.
- 28 Moseley AM, Elkins MR, Herbert RD, et al. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol*. 2009;62:1021-1030.
- 29 *Position Statement: Standards of Physical Therapy Practice*. London, United Kingdom: World Confederation for Physical Therapy; 2011.
- 30 Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiology Community Health*. 2004;58: 635-641.
- 31 Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51-68.
- 32 Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials*. 2011;12:58.
- 33 Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ*. 2010;341:33.
- 34 Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol*. 2005;34:79-87.
- 35 Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Phys Ther Rev*. 2009;14:36-49.
- 36 Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*. 2002;287:2801-2804.
- 37 Soares HP, Daniels S, Kumar A, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ*. 2004;328:22-24.
- 38 MacCallum R. Model specification, procedures, strategies, and related issues. In: Hoyle RH, ed. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications; 1995:16-36.