

PROTOCOL

Open Access

Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a study protocol

Bruno R da Costa^{1*}, Nina M Resta¹, Brooke Beckett¹, Nicholas Israel-Stahre¹, Alison Diaz¹, Bradley C Johnston^{2,3,4}, Matthias Egger⁵, Peter Jüni^{5,6} and Susan Armijo-Olivo^{7,8}

Abstract

Background: The Cochrane risk of bias (RoB) tool has been widely embraced by the systematic review community, but several studies have reported that its reliability is low. We aim to investigate whether training of raters, including objective and standardized instructions on how to assess risk of bias, can improve the reliability of this tool. We describe the methods that will be used in this investigation and present an intensive standardized training package for risk of bias assessment that could be used by contributors to the Cochrane Collaboration and other reviewers.

Methods/Design: This is a pilot study. We will first perform a systematic literature review to identify randomized clinical trials (RCTs) that will be used for risk of bias assessment. Using the identified RCTs, we will then do a randomized experiment, where raters will be allocated to two different training schemes: minimal training and intensive standardized training. We will calculate the chance-corrected weighted Kappa with 95% confidence intervals to quantify within- and between-group Kappa agreement for each of the domains of the risk of bias tool. To calculate between-group Kappa agreement, we will use risk of bias assessments from pairs of raters after resolution of disagreements. Between-group Kappa agreement will quantify the agreement between the risk of bias assessment of raters in the training groups and the risk of bias assessment of experienced raters. To compare agreement of raters under different training conditions, we will calculate differences between Kappa values with 95% confidence intervals.

Discussion: This study will investigate whether the reliability of the risk of bias tool can be improved by training raters using standardized instructions for risk of bias assessment. One group of inexperienced raters will receive intensive training on risk of bias assessment and the other will receive minimal training. By including a control group with minimal training, we will attempt to mimic what many review authors commonly have to do, that is—conduct risk of bias assessment in RCTs without much formal training or standardized instructions. If our results indicate that an intense standardized training does improve the reliability of the RoB tool, our study is likely to help improve the quality of risk of bias assessments, which is a central component of evidence synthesis.

Keywords: Systematic review, Meta-analysis, Risk of bias, RCT, Cochrane

* Correspondence: bdacosta@fiu.edu

¹Department of Physical Therapy, Florida International University, AHC3-430
11200 8 St, Miami, FL, USA

Full list of author information is available at the end of the article

Background

Systematic reviews and meta-analyses of randomized clinical trials (RCTs) are central to evidence-based clinical decision-making [1,2]. RCTs are considered the gold standard design when assessing the effectiveness of treatment interventions. Appropriately conducted RCTs may eliminate confounding, allowing decision-makers to infer that changes observed in the outcome of interest are causally linked with the experimental intervention.

If results of RCTs included in a meta-analysis are biased, so will the results of the meta-analysis [3,4]. To address this problem, it is recommended that the risk of bias in RCTs is taken into consideration when conducting meta-analysis. A method commonly used for this purpose involves the stratification of meta-analyses according to RCTs with low or high risk of bias.

In 2008, the Cochrane Collaboration published a tool and guidelines for the assessment of risk of bias in RCTs [5,6]. The risk of bias tool has been widely embraced by the systematic review community [7]. The items in this tool address six domains of bias, which are classified as low, high, or unclear risk of bias. The selection of domains of bias was based on empirical evidence and theoretical considerations, focusing on methodological issues that are likely to influence the results of RCTs.

Several studies reported that the reliability of the risk of bias tool is low [8-10]. Reliability of the risk of bias tool can be assessed between two raters of the same research group, when for instance, they assess the risk of bias of RCTs included in a meta-analysis in duplicate. It can also be assessed across research groups, if the risk of bias was assessed for a trial included in two different meta-analyses by two different research groups. Disagreements between two raters of the same research group may be less problematic since they will normally discuss their ratings to come to a consensus. Disagreements between raters from different research groups will be more problematic, for example, if for the same outcome a trial is considered at low risk of bias in one meta-analysis, but is at high risk of bias in another one. Low reliability of risk of bias assessments can then ultimately have repercussions on decision-making and quality of patient care [11,12].

We recently found that the reliability of the risk of bias tool might be improved by intensive standardized training of raters [8]. However, to our knowledge, no formal evaluation of such a training intervention has been performed. We therefore aim to investigate whether training of raters, with objective and standardized instructions on how to assess risk of bias, can improve the within and between pairs of rater reliability of the Cochrane RoB tool. Here, we describe the methods and the intensive standardized training package for risk of bias assessment that will be used in this study.

Methods/Design

Study design

This is a pilot study. The first component is a systematic literature review to identify RCTs that will be used for risk of bias assessment. Using the identified RCTs, the second component of our investigation is a randomized experiment, where raters will be allocated to two different levels of training on risk of bias assessment: minimal training and intensive standardized training.

Literature search and trial selection

We will search PubMed from inception using database-specific search strategy (Figure 1). We will include every randomized or quasi-randomized clinical trial in patients with knee osteoarthritis that compared a physical therapy intervention to another physical therapy intervention, sham intervention, or no treatment, which assessed patient-reported pain. The following physical therapy interventions will be considered: land-based exercise, aquatic exercise, manual therapy, electric stimulation therapy, and diathermy. We will only consider studies published in English. No further restrictions will be applied. Two raters will screen reports for eligibility independently in duplicate. Disagreements will be resolved by a senior author (BdC).

Data extraction

We will use standardized, piloted data extraction forms to extract information on publication year, sample size, and type of intervention. We will assess risk of bias for selected items of the risk of bias tool, namely sequence of generation, allocation concealment, blinding (participants, personnel, and assessors), and incomplete outcome data. Although a potentially important source of bias, we will not assess selective outcome reporting in our study due to feasibility issues with such assessment [7]. Within pairs of raters, data extraction will be conducted independently and in duplicate. Potential disagreements within pairs of raters will be solved by discussion until consensus is reached.

Training on risk of bias assessment

Six raters will assess the risk of bias of every included trial. Four of these raters are doctoral students of physical therapy without previous experience in risk of bias assessment, and two raters are experienced risk of bias assessors. The experienced raters have each been involved in over 15 systematic reviews that included methodological quality assessment. We will use simple randomization (computer-generated numbers <http://www.randomizer.org/form.htm>) to allocate two students to minimal training and two to intensive training. Randomization will be performed remotely by one of the authors (SAO), who had no contact with the students. Students will not be informed to which

| Search Strategy* | |
|------------------|---|
| Step | Search Terms |
| 1 | Osteoarthritis [Title]. |
| 2 | Knee [Title/Abstract]. |
| 3 | AND/1-2 |
| 4 | Exercise therapy [Mesh]. |
| 5 | Electric stimulation therapy [Mesh]. |
| 6 | Hydrotherapy [Mesh]. |
| 7 | Diathermy [Mesh]. |
| 8 | Musculoskeletal manipulations [Mesh]. |
| 9 | OR/4-8 |
| 10 | Pain [Title/Abstract]. |
| 11 | Randomized controlled trial [Publication Type]. |
| 12 | English [Language]. |
| 13 | 3 AND 9 AND 10 AND 11 AND 12 |
| 14 | Surgery[All Fields]. OR Arthroplasty[All Fields]. OR Protocol[All Fields]. OR Pilot[All Fields]. OR "Drug therapy"[Mesh]. |
| 15 | 13 NOT 14 |

Figure 1 Search strategy. *Search strategy conducted on PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). The search was last updated on 20 March 2014.

training group they have been randomized, and they will be instructed not to discuss their training with each other to minimize the risk of 'contamination' [5]. After data extraction is completed, we will ask students to guess in which group they were allocated, whether there was any event during data extraction that made them aware of their group allocation, and if this affected their performance in this study.

Minimal training will mimic assessments made by raters without formal training. It will consist of a single lecture for about 60 min on the definition and importance of each of the assessed domains of bias, without specific or standardized instructions on how to conduct the assessment. In this lecture, the raters will learn how the biases could occur in an RCT (for example, staff recruiting patients who are not concealed to allocation may tamper with random allocation sequences), and they will be shown results of empirical investigations that observed the effect of these biases in clinical treatment effects. The raters will be provided with an article, as an optional reading material, that describes the risk of bias tool [6] as well as chapter 8 of the Cochrane Handbook for Systematic Reviews of Interventions [5], which specifically addresses the assessment of risk of bias of studies included in a systematic review.

Raters allocated to intensive training will receive the same lecture for about 60 min. In addition, they will

receive standardized instructions on how to assess each of the domains (Additional file 1). The standardized instructions were based on the Cochrane Handbook [5] and adapted as deemed necessary to increase their objectivity and thus minimize misinterpretations for the assessment of trials of physical therapy in patients with knee osteoarthritis. One of the experienced raters (BdC) will discuss these instructions with them and will give them the opportunity to clarify any questions they may have. This rater is an experienced clinical epidemiologist who was trained by one of the authors of the risk of bias tool (PJ) and has been involved in over 20 systematic reviews that included methodological quality assessment. Next, raters will assess risk of bias in a purposively selected sample of ten articles, which will not be part of the final study sample. One of the experienced raters (BdC) will discuss their assessments after five and ten training articles have been assessed. During these sessions, disagreements between the assessment conducted by the experienced rater and the other raters will be identified. The experienced rater will then provide a rationale to justify his assessment, and discussion will take place until all questions are clarified. The assessments of the raters allocated to intensive training will thus be calibrated with the assessments of the experienced rater.

The raters in both groups will be instructed to not discuss their risk of bias assessment with others.

The study protocol was approved by the research ethics committee of the Florida International University (IRB-14-0110). We will obtain written informed consent from each student rater. This study was not registered in PROSPERO.

Analysis

We will tabulate the characteristics of included trials and the risk of bias assessments of the three groups of raters (minimal training, intensive training, and experienced raters); before and after consensus. We will then calculate the chance-corrected weighted Kappa with 95% confidence intervals to quantify agreement within and between the three groups of raters for each of the domains of the risk of bias tool. To calculate between-group Kappa agreement, we will use risk of bias assessments from pairs of raters after resolution of disagreements. Kappa values range from 0 to 1, with higher values indicating higher agreement between raters. To calculate the weighted Kappa agreement, risk of bias classification will be ordered as follows: low, unclear, and high risk of bias. Criteria proposed by Byrt [13] will be used to interpret Kappa values. Values between 0.93 and 1.00 represented excellent agreement; 0.81 and 0.92 very good agreement; 0.61 and 0.80 good agreement; 0.41 and 0.60 fair agreement; 0.21 and 0.40 slight agreement, 0.01 and 0.20 poor agreement; and 0.00 or less will be considered to reflect no agreement.

The main outcome of our study is the comparison of the accuracy of assessment across groups, that is, a comparison of the agreement between raters receiving intensive training and experienced raters with the agreement between raters receiving minimal training and experienced raters. To compare agreement of raters under different training conditions, we will calculate the differences between Kappa values for within- and between-group agreement. To address deviations from normal distribution, we will bootstrap the difference in Kappa values using bias correction and acceleration to derive 95% confidence intervals and *P*-values [14]. Please see Additional file 2 for assumptions used for the power analysis.

To explore whether quality of reporting influences agreement, we will stratify the analysis according to publication date (before the first CONSORT statement revision in 2001 vs 2001 [15] and later; and before the latest CONSORT statement revision in 2010 [16] vs 2010 and later), assuming that reporting quality of RCTs in physical therapy improved after the publication of the CONSORT statement [17,18]. To investigate whether methodological quality influences agreement between raters, we will stratify the analysis by trial size (<100 and ≥100 patients randomized per trial arm), assuming that trial size is associated with methodological quality [19]. A sensitivity analysis will be conducted on larger (≥100 patients

randomized per trial arm) trials published after 2010 (i.e. after the latest revision of the CONSORT statement was published). All *P*-values will be two-sided. Analysis will be conducted in STATA, release 13 (StataCorp, College Station, TX).

Discussion

This study will investigate whether the reliability of the risk of bias tool can be improved by training raters using standardized instructions for risk of bias assessment. We will calculate Kappa coefficients to quantify the agreement between experienced raters and inexperienced raters. Two inexperienced raters will receive intensive training on risk of bias assessment and the other two will receive minimal training. By including a control group with minimal training, we will attempt to mimic what many review authors commonly have to do, that is—conduct risk of bias assessments of RCTs without any formal training or standardized instructions. If our results indicate that an intensive standardized training does improve the reliability of the risk of bias tool, our study is likely to help improve the quality of risk of bias assessment, which is a central component of evidence synthesis [4]. The standardized training package for risk of bias assessment could then be used by organizations such as the Cochrane Collaboration and the systematic review community at large. By publishing this protocol, we would also like to make the objectives and pre-specified study design transparent to readers, as has been recommended for methodological studies [20].

Strengths and limitations

Our study had two major strengths. First, we will include raters completely inexperienced with the risk of bias assessment to investigate the effect of intensive training on the reliability of the risk of bias tool. By including inexperienced raters, we believe we will be more likely to observe an effect of intensive training, if one indeed exists. If raters were already experienced with the risk of bias assessment, there could be limited room for improvement as postulated in a previous study that investigated the effect of training on a similar method for methodological quality assessment [21]. Second, raters will be randomly allocated to training groups, and central randomization will be performed to conceal the random sequence of allocation. We acknowledge that this approach may not create a balanced distribution of potential confounders across the two training groups due to the low number of individuals randomized, but the group of students randomized will be fairly homogeneous and confounding therefore unlikely.

Our results can potentially be influenced by performance bias. If raters in the control group understand that they are not receiving the best training available in our study, they may feel discouraged to try and perform risk

of bias assessments as best as they can. This could in turn lead to an artificially lower reliability of the risk of bias tool with minimal training as compared to intensive training. Alternatively, they could seek additional training elsewhere or be prompted to self-study. To try and minimize the risk of such performance bias, raters will not be informed to which training group they have been randomized, and they will be instructed to not discuss with each other any characteristics of their training. The use of minimal training as a control intervention may lead to an underestimation of the effect of our intensive training, however. Although 'no training' could be used as a control intervention instead of minimal training, this could substantially increase the risk of performance bias in our study, as explained above.

Although the standardized instructions were primarily based on the Cochrane Handbook instructions for risk of bias assessment, we added or adapted some of the criteria to facilitate raters' decision-making in the present investigation. Some of these criteria are not solely based on empirical evidence but also on our own experience with analyses of individual RCTs and meta-analyses of RCTs (Additional file 1). The main example concerns the assessment of incomplete outcome data. To facilitate decision-making, we used thresholds of drop-out rates to define low and high risk of bias for assessing incomplete data. These thresholds were chosen based on evidence that this drop-out rate could be linked to biases [22,23]. Moreover, some of these adaptations may not apply to systematic reviews with different research questions. For instance, because outcomes in our sample of physical therapy trials could potentially be influenced by at least some degree of performance bias, blinding of patients and therapists were by definition deemed necessary. Conversely, the Cochrane Handbook recommends that raters first assess the need for patient and therapist blinding in each trial in light of the outcome of interest and to take this information into consideration when assessing the risk of performance bias. Although we fully agree with this recommendation, it was not included in our standardized instructions for the reason explained above.

The low number of raters randomized to intervention groups will be a limitation for the generalizability of our findings. Due to feasibility issues, we included the minimal number of participants needed to calculate Kappa agreements within each study condition. If results of the present investigation indicate that intensive training may indeed improve the reliability of the risk of bias tool, a future study including a larger number of raters could replicate our methods to address this limitation. Generalizability may be further hindered by the characteristics of the trials assessed in our study and of the individual providing the training on risk of bias assessment. Accuracy of the risk of

bias assessment could vary if trials with different patient populations, interventions, and outcomes were assessed. Likewise, the quality of the training on the risk of bias assessment may vary according to the skills of the person who provides the training and the method used to deliver such training.

Previous research

The risk of bias tool has been extensively used in many Cochrane reviews, albeit the information of the inter-rater reliability of the risk of bias tool is rather limited. To date, five studies [9,10,24-26] have investigated the inter-rater reliability of the risk of bias tool, but none have proposed and investigated ways to improve its reliability. The inter-rater agreement for the individual domains of the risk of bias tool has been found to range from poor (Kappa = 0.13 for selective reporting) to substantial (Kappa = 0.74 for sequence generation) [25]. Our team recently investigated the between-group Kappa agreement of the risk of bias tool comparing ratings from Cochrane reviewers to ratings from our team [8]. Most of our Kappa values were considerably higher than those reported in previous studies. We suspect two key factors may explain these differences. Although we used the Cochrane Handbook guidelines for risk of bias assessments, we first predefined specific decision rules to assess the individual domains of the tool. Second, we used an intensive training for the raters to standardize the decision-making process. This could partly explain the results of our previous study. The current study will investigate this hypothesis.

Additional files

Additional file 1: Guidelines for evaluating the risk of bias in physical therapy trials with patients with knee osteoarthritis.

Additional file 2: Assumptions of the power analysis.

Abbreviations

RoB: risk of bias; *RCT*: randomized clinical trial; *PROSPERO*: international prospective register of systematic reviews; *CONSORT*: consolidated standards of reporting trials.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BdC conceived the study, developed a first study design proposal, and wrote the first draft of the manuscript. NMR, BB, NIS, and AD provided feedback about the study design and read and approved the final manuscript. BCJ, ME, PJ, and SAO helped develop a first study design proposal, provided feedback about the study design, and read and approved the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the Nicole Wertheim College of Nursing and Health Sciences, Florida International University, Miami, for providing funding for the publication fees of this paper.

Author details

¹Department of Physical Therapy, Florida International University, AHC3-430 11200 8 St, Miami, FL, USA. ²Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada. ³Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. ⁴Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada. ⁵Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland. ⁶Department of Clinical Research, CTU Bern, University of Bern, Bern, Switzerland. ⁷CLEAR (Connecting Leadership and Research) Outcomes Research Program, Faculty of Nursing, University of Alberta, Edmonton, AB, Canada. ⁸Faculty of Rehabilitation Medicine, Department of Physical Therapy, University of Alberta, Edmonton, AB, Canada.

Received: 15 September 2014 Accepted: 8 December 2014

Published: 13 December 2014

References

1. Cook DJ, Mulrow CD, Haynes RB: **Systematic reviews: synthesis of best evidence for clinical decisions.** *Ann Intern Med* 1997, **126**:376–380.
2. Egger M, Smith GD: **Meta-analysis: potentials and promise.** *BMJ* 1997, **315**:1371–1374.
3. Egger M, Smith GD, Sterne JA: **Uses and abuses of meta-analysis.** *Clin Med* 2001, **1**:478–484.
4. Juni P, Altman DG, Egger M: **Systematic reviews in health care: assessing the quality of controlled clinical trials.** *BMJ* 2001, **323**:42–46.
5. Higgins J, Altman D: **"Chapter 8: Assessing risk of bias in included studies"**. In *Cochrane Handbook for Systematic Reviews of Interventions version 5.0*. Edited by Higgins J, Green S. Chichester, UK: John Wiley & Sons, Ltd; 2008.
6. Higgins JPT, Altman DG, Goetzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA: **The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.** *BMJ* 2011, **343**:d5928. doi: 10.1136/bmj.d5928.
7. Sterne JA: **Why the Cochrane risk of bias tool should not include funding source as a standard item.** *Cochrane Database Syst Rev* 2013, **12**:ED000076.
8. Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes CJ, Ha C, Cummings GG: **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**:1–10.
9. Graham N, Haines T, Goldsmith CH, Gross A, Burnie S, Shahzad U, Talovikova E: **Reliability of three assessment tools used to evaluate randomized controlled trials for treatment of neck pain.** *Spine* 2011, **37**:515–522.
10. Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM: **Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs.** *J Clin Epidemiol* 2012, **66**:973–981.
11. da Costa BR, Hilfiker R, Egger M: **PEDro's bias: summary quality scores should not be used in meta-analysis.** *J Clin Epidemiol* 2013, **66**:75–77.
12. Juni P, Witschi A, Bloch R, Egger M: **The hazards of scoring the quality of clinical trials for meta-analysis.** *Jama* 1999, **282**:1054–1060.
13. Byrt T: **How good is that agreement?** *Epidemiology (Cambridge, Mass)* 1996, **7**:561–1996.
14. Efron B: **Better bootstrap confidence intervals.** *J Am Stat Assoc* 1987, **82**:171–185.
15. Moher D, Schulz KF, Altman D: **The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials.** *JAMA* 2001, **285**:1987–1991.
16. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: **CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.** *J Clin Epidemiol* 2010, **63**:e1–e37.
17. To MJ, Jones J, Emara M, Jadad AR: **Are reports of randomized controlled trials improving over time? A systematic review of 284 articles published in high-impact general and specialized medical journals.** *PLoS One* 2013, **8**:e84779.
18. Moher D, Jones A, Lepage L: **Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation.** *JAMA* 2001, **285**:1992–1995.
19. Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, Egger M, Jüni P: **Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study.** *BMJ* 2010, **341**:c3515.
20. Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, Akl EA, Sun X, Johnston BC, Briel M, Busse JW, Glujovsky D, Granados CE, Iorio A, Irfan A, García LM, Mustafa RA, Ramirez-Morera A, Solà I, Tikkinen KA, Ebrahim S, Vandvik PO, Zhang Y, Selva A, Sanabria AJ, Zazueta OE, Vernooij RW, Schünemann HJ, Guyatt GH: **A methodological survey of the analysis, reporting and interpretation of Absolute Risk Reduction in systematic reviews (ARROW): a study protocol.** *Syst Rev* 2013, **2**:113.
21. Fourcade L, Boutron I, Moher D, Ronceray L, Baron G, Ravaud P: **Development and evaluation of a pedagogical tool to improve understanding of a quality checklist: a randomised controlled trial.** *PLoS Clin Trials* 2007, **2**:e22.
22. Unnebrink K, Windeler J: **Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases.** *Stat Med* 2001, **20**:3931–3946.
23. Wright CC, Sim J: **Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis.** *J Clin Epidemiol* 2003, **56**:833–842.
24. Hartling L, Bond K, Vandermeer B, Seida J, Dryden D, Rowe B: **Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma.** *PLoS Med* 2011, **6**:1–6.
25. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Seida JK, Klassen TP: **Risk of bias versus quality assessment of randomised controlled trials: cross sectional study.** *BMJ* 2009, **339**:1017.
26. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG: **Assessment of study quality for systematic reviews: a comparison of the Cochrane collaboration risk of bias tool and the effective public health practice project quality assessment tool: methodological research.** *J Eval Clin Pract* 2012, **18**:12–18.

doi:10.1186/2046-4053-3-144

Cite this article as: da Costa et al.: Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a study protocol. *Systematic Reviews* 2014 **3**:144.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

