

ORIGINAL ARTICLE

## Quality of systematic reviews: an example of studies comparing artificial disc replacement with fusion in the cervical spine

Osama A. Tashani<sup>1\*</sup>, Hanan El-Tumi<sup>1</sup> and Khaled Aneiba<sup>1,2</sup>

<sup>1</sup>Centre for Pain Research, Faculty of Health and Social Sciences, Leeds Beckett University, Leeds, UK;

<sup>2</sup>Orthopaedics Department, University Hospital of North Tees, Stockton-on-Tees, UK

Cervical artificial disc replacement (C-ADR) is now an alternative to anterior cervical discectomy and fusion (ACDF). Many studies have evaluated the efficacy of C-ADR compared with ACDF. This led to a series of systematic reviews and meta-analyses to evaluate the evidence of the superiority of one intervention against the other. The aim of the study presented here was to evaluate the quality of these reviews and meta-analyses. Medline via Ovid, Embase, and Cochrane Library were searched using the keywords: (total disk replacement, prosthesis, implantation, discectomy, and arthroplasty) AND (cervical vertebrae, cervical spine, and spine) AND (systematic reviews, reviews, and meta-analysis). Screening and data extraction were conducted by two reviewers independently. Two reviewers then assessed the quality of the selected reviews and meta-analysis using 11-item AMSTAR score which is a validated measurement tool to assess the methodological quality of systematic reviews. Screening of full reports of 46 relevant abstracts resulted in the selection of 15 systematic reviews and/or meta-analyses as eligible for this study. The two reviewers' inter-rater agreement level was high as indicated by kappa of  $>0.72$ . The AMSTAR score of the reviews ranged from 3 to 11. Only one study (a Cochrane review) scored 100% (AMSTAR 11). Five studies scored below (AMSTAR 5) indicating low-quality reviews. The most significant drawbacks of reviews of a score below 5 were not using an extensive search strategy, failure to use the scientific quality of the included studies appropriately in formulating a conclusion, not assessing publication bias, and not reporting the excluded studies. With a significant exception of a Cochrane review, the methodological quality of systematic reviews evaluating the evidence of C-ADR versus ACDF has to be improved.

Keywords: *cervical spine; arthroplasty; systematic reviews; surgical interventions; fusion*

\*Correspondence to: Osama A. Tashani, Centre for Pain Research, Faculty of Health and Social Sciences, Leeds Beckett University, Portland PD609, City Campus, Leeds LS1 3HE, UK, Email: O.Tashani@leedsbeckett.ac.uk

Received: 15 June 2015; Accepted in revised form: 28 June 2015; Published: 22 July 2015

A systematic review is an attempt to collect 'all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question' (Cochrane Handbook for Systematic Reviews of interventions, handbook.cochrane.org, Section 1.2.2). If the results of the individual studies are combined to produce an overall statistic, this is usually called a meta-analysis. Systematic reviews and meta-analyses inform medical and health practitioners and are highly regarded as informative and authoritative in informing decision-makers who formulate the guidelines for practice. Unfortunately, not all systematic reviews and meta-analyses follow the robust methodology of systematic reviewing that are publicised by Cochrane Collaboration and other

organisations, Universities and Research Centres. It is essential to implement scientific methods in evaluating the evidence available in the clinical trials and other studies that are reviewed and this should be complemented by transparent reporting of the methods used. There is a need to alert clinicians and systematic reviewers to the indispensable quality requirement of systematic reviews and specifically in evaluating interventions in which randomisation of patients is challenging in clinical trials. Such interventions that are very difficult to randomise are surgical procedures. In this study, quality of systematic reviews of surgical intervention at the cervical spine will be assessed to provide an example of problems associated with systematic reviewing and meta-analysis of these interventions.

Cervical artificial disc replacement (C-ADR) is now an alternative to anterior cervical discectomy and fusion (ACDF). Some biomechanical studies claim that C-ADR is significantly superior than ACDF in terms of restoring sagittal profile particularly cervical lordosis as well as offering a better range of motion in the caudal segment compared to cranial segment (1). Many other studies have evaluated the efficacy of C-ADR compared to ACDF in a range of clinical and patient outcomes, including pain, quality of life, and fusion rate. This led to a series of systematic reviews and meta-analyses to evaluate the evidence of the superiority of one intervention over the other. Therefore, the aim of the study presented here was to evaluate the quality of these reviews and meta-analyses using an up-to-date assessment tool.

## Methods

Medline via Ovid, Embase, and Cochrane Library were searched using the keywords: (total disk replacement, prosthesis, implantation, discectomy, and arthroplasty) AND (cervical vertebrae, cervical spine, and spine) AND (systematic reviews, reviews, and meta-analysis). The initial search was conducted on 18 August 2013 and updated on 02 February 2015. Eligibility criteria, which were applied by two reviewers independently (OAT, KA), were: the article should be a systematic review and/or meta-analysis of randomised controlled trials (RCTs) or cohort studies that compared C-ADR with ACDF at the cervical region for one or two levels and for follow-up period of at least more than 6 months and should be published in a peer-reviewed journal. No time or language limits were applied in the selection process. Two reviewers (OAT, HE) then assessed the quality of the selected reviews and meta-analyses using the 11-item AMSTAR score which is a validated measurement tool to assess the methodological quality of systematic reviews (2). The agreement between the two reviewers was tested using the fixed-margin kappa. The 11 items are summarised as follows:

1. The research question and inclusion criteria should be established before the conduct of the review.
2. At least two reviewers should extract the data independently and a consensus procedure to resolve their disagreement should be in place.
3. At least two electronic sources should be searched.
4. The authors should state that they searched for reports regardless of their publication type.
5. A list of included and excluded studies should be provided.
6. The characteristics of the included studies should be provided.
7. The scientific quality of the included studies should be assessed and documented.
8. The scientific quality of the included studies should be used appropriately in formulating conclusions.
9. The methods used to combine the findings of studies should be appropriate (Need to assess homogeneity).
10. The likelihood of publication bias should be assessed.
11. The conflict of interest should be declared and included.

The two assessors had to decide whether the systematic review or the meta-analysis meets the criterion (score 1) or cannot decide on an item or that the review does not meet the criterion (score 0).

## Results

The initial search produced 636 citations of which 49 were duplicates. Screening of full reports of 46 relevant abstracts resulted in the selection of 15 eligible systematic reviews and/or meta-analyses. All selected articles were published in English. The two reviewers' inter-rater agreement level was high, as indicated by kappa of  $>0.72$  (3). In other words, the two assessors reached perfect agreement in 12 out of the 15 systematic reviews assessed. The AMSTAR score of the reviews ranged from 3 to 11 (see Table 1). Only one study (a Cochrane review) scored 100% (AMSTAR 11). Five studies scored below 5 (AMSTAR) indicating that these reviews are either of low quality or failed to report their methodology in selecting and critically appraising the studies included. The most significant drawbacks of reviews of a score below 5 were: not using an extensive search strategy, failure to use the scientific quality of the included studies appropriately in formulating a conclusion, not assessing publication bias, and not reporting the excluded studies. The AMSTAR items that were consistently missing were: 5 (86.7% of the reviews did not report excluded studies), 8 (66.7% of the reviews failed to account for the quality of articles when formulating conclusions), 10 (80% of the reviews did not attempt to measure publication bias), and 11 (80% of authors failed to disclose or report any conflict of interest).

## Discussion and conclusion

With the significant exception of Cochrane reviews, the methodological quality of systematic reviews evaluating the evidence of C-ADR versus ACDF has to be improved specifically in transparent reporting. It is clear that conflict of interest disclosure might have been disclosed to journal editors but failure to report this in the review will lead to lower score in AMSTAR and the readers might question the bias of authors. More significantly, reporting of excluded studies, with reasons of exclusion, in systematic reviews is an essential part of the reviewing process, and authors of systematic reviews and meta-analyses should be encouraged to report this as this assures the readers and the users of the results of the systematic reviews that the article selection process is robust and unbiased. There are, of course, some challenges in the RCTs of surgically

Table 1. Quality assessment of included systematic reviews

Ref.	Year	Assessor	AMSTAR item (see text)											Total score	No. of agreements	Mean score	
			1	2	3	4	5	6	7	8	9	10	11				
9	2012	A	1	1	1	1	1	1	1	1	1	1	1	1	11	11	11
		B	1	1	1	1	1	1	1	1	1	1	1	1	11		
18	2012	A	1	1	1	1	0	1	1	1	1	1	1	1	10	11	10
		B	1	1	1	1	0	1	1	1	1	1	1	1	10		
11	2012	A	1	1	1	1	1	1	1	0	1	1	0	9	11	9	
		B	1	1	1	1	1	1	1	0	1	1	0	9			
17	2010	A	1	1	1	1	0	1	1	0	1	1	1	9	11	9	
		B	1	1	1	1	0	1	1	0	1	1	1	9			
7	2013	A	1	1	1	1	0	1	1	0	1	0	0	7	11	7	
		B	1	1	1	1	0	1	1	0	1	0	0	7			
16	2013	A	1	1	1	1	0	1	1	0	0	0	1	7	7	7	
		B	1	1	1	0	0	1	0	0	1	1	1	7			
14	2012	A	1	1	1	0	0	1	1	1	1	0	0	7	11	7	
		B	1	1	1	0	0	1	1	1	1	0	0	7			
6	2011	A	1	1	1	1	0	1	1	0	1	0	0	7	11	7	
		B	1	1	1	1	0	1	1	0	1	0	0	7			
13	2011	A	1	1	1	1	0	1	1	1	0	0	0	7	11	7	
		B	1	1	1	1	0	1	1	1	0	0	0	7			
4	2010	A	1	1	1	1	0	1	1	0	0	0	0	6	11	6	
		B	1	1	1	1	0	1	1	0	0	0	0	6			
15	2012	A	1	1	1	0	0	1	1	1	1	0	0	7	9	6	
		B	1	1	0	0	0	1	1	0	1	0	0	5			
5	2012	A	1	1	0	1	0	1	1	0	0	0	0	5	11	5	
		B	1	1	0	1	0	1	1	0	0	0	0	5			
8	2012	A	1	1	0	1	0	1	0	0	1	0	0	5	11	5	
		B	1	1	0	1	0	1	0	0	1	0	0	5			
10	2012	A	1	0	0	1	0	1	0	0	1	0	0	4	11	4	
		B	1	0	0	1	0	1	0	0	1	0	0	4			
12	2013	A	1	0	1	0	0	0	0	0	1	0	0	3	9	2	
		B	1	0	0	0	0	0	0	0	0	0	0	1			

Percent of overall agreement of the two assessors was: 0.799999. Fixed-marginal kappa: 0.74.

interventions including the limitation to select a random sample of patients, the impracticality of blinding surgical interventions. Nevertheless, systematic reviews have to apply rigorous methodology to inform clinicians, patients, and guidelines decision bodies.

### Recommendations

It is essential to follow the guidelines of Cochrane Collaboration, or an equivalent organisation such as the Centre for Reviews and Dissemination at York, when conducting systematic reviews or meta-analyses. There are many resources available for training and consulting whenever a research team plans to do a systematic review. It must be acknowledged that the systematic review is similar to any other scientific articles and can be of varying quality. Systematic reviewing is team work and a team of three or more is usually required with at least one statistician included in the team. The team has to have

an expertise in the topic reviewed and has to carry the systematic review according to a pre-planned protocol. The work usually takes 9–12 months to complete depending on the number of articles retrieved. The Centre for Reviews and Dissemination at York summarise the steps to conduct a systematic review as follows:

1. Clear stated objectives.
2. Pre-defined eligibility criteria for studies.
3. A systematic search of literature with at least two databases searched.
4. Assessment of validity of findings (e.g. risk of bias).
5. Systematic presentation and synthesis of evidence.

We also recommend the use of PRISMA checklist ([www.prisma-statement.org/2.1.2 - PRISMA 2009 Checklist.pdf](http://www.prisma-statement.org/2.1.2-PRISMA%202009%20Checklist.pdf)) and AMSTAR assessment tool ([www.amstar.ca/Amstar\\_Checklist.php](http://www.amstar.ca/Amstar_Checklist.php)) when conducting systematic reviews.

## Authors' contributions

The first and the last authors developed the concept of this manuscript. All authors contributed to the writing of the first draft and all approved the last draft.

## Conflict of interest and funding

This research was financially supported by Trimedica Ltd and Alphatec Spine. The two sponsors had no role in the study, decision to publish, or any copyright relevant to the study.

## References

1. Anderson PA, Rouleau JP. Intervertebral disc arthroplasty. *Spine*. 2004; 29: 2779–86.
2. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007; 7: 10.
3. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012; 22: 276–82.
4. Zechmeister I, Winkler R, Mad P. Artificial total disc replacement versus fusion for the cervical spine: a systematic review. *Eur Spine J*. 2011; 20: 177–84.
5. Mummaneni PV, Burkus JK, Haid RW, Traynelis VC, Zdeblick TA. Clinical and radiographic analysis of cervical disc arthroplasty compared with allograft fusion: a randomized controlled clinical trial. *J Neurosurg Spine*. 2007; 6: 198–209.
6. Yu L, Song Y, Yang X, Lv C. Systematic review and meta-analysis of randomized controlled trials: comparison of total disk replacement with anterior cervical decompression and fusion. *Orthopedics*. 2011; 34: 772.
7. Yin S, Yu X, Zhou S, Yin Z, Qiu Y. Is cervical disc arthroplasty superior to fusion for treatment of symptomatic cervical disc disease? A meta-analysis. *Clin Orthop Relat Res*. 2013; 471: 1904–19.
8. Upadhyaya CD, Wu J-C, Trost G, Haid RW, Traynelis VC, Tay B, et al. Analysis of the three United States Food and Drug Administration investigational device exemption cervical arthroplasty trials: clinical article. *J Neurosurg Spine*. 2012; 16: 216–28.
9. Boselie TF, Willems PC, van Mameren H, de Bie R, Benzel EC, van Santbrink H. Arthroplasty versus fusion in single-level cervical degenerative disc disease. *Cochrane Database Syst Rev* 2012; 9: CD009173.
10. McAfee PC, Reah C, Gilder K, Eisermann L, Cunningham B. A meta-analysis of comparative outcomes following cervical arthroplasty or anterior cervical fusion: results from 4 prospective multicenter randomized clinical trials and up to 1226 patients. *Spine (Phila Pa 1976)* 2012; 37: 943–52.
11. Yang B, Li H, Zhang T, He X, Xu S. The incidence of adjacent segment degeneration after cervical disc arthroplasty (CDA): a meta analysis of randomized controlled trials. *PLoS One* 2012; 7: e35032.
12. Verma K, Gandhi SD, Maltenfort M, Albert TJ, Hilibrand AS, Vaccaro AR, et al. Rate of adjacent segment disease in cervical disc arthroplasty versus single-level fusion: meta-analysis of prospective studies. *Spine*. 2013; 38: 2253–7.
13. Cepoiu-Martin MMDM, Faris PP, Lorenzetti DM, Prefontaine EB, Noseworthy TMDM, Sutherland LMDM. Artificial cervical disc arthroplasty: a systematic review. *Spine*. 2011; 36: E1623–33.
14. Gebremariam L, Koes BW, Peul WC, Huisstede BM. Evaluation of treatment effectiveness for the herniated cervical disc: a systematic review. *Spine (Phila Pa 1976)* 2012; 37: E109–18.
15. Anderson PA, Sasso RC, Hipp J, Norvell DC, Raich A, Hashimoto R. Kinematics of the cervical adjacent segments after disc arthroplasty compared with anterior discectomy and fusion: a systematic review and meta-analysis. *Spine*. 2012; 37: S85–95.
16. Gao Y, Liu M, Li T, Huang F, Tang T, Xiang Z. A meta-analysis comparing the results of cervical disc arthroplasty with anterior cervical discectomy and fusion (ACDF) for the treatment of symptomatic cervical disc disease. *J Bone Joint Surg*. 2013; 95: 555–61.
17. Bartels RHM, Donk R, Verbeek ALM. No justification for cervical disk prostheses in clinical practice: a meta-analysis of randomized controlled trials. *Neurosurgery*. 2010; 66: 1153–60.
18. Fallah A, Akl EA, Ebrahim S, Ibrahim GM, Mansouri A, Foote CJ, et al. Anterior cervical discectomy with arthroplasty versus arthrodesis for single-level cervical spondylosis: a systematic review and meta-analysis. *PLoS One* 2012; 7: e43407.