AO SPINE

Critically Low Confidence in the Results Produced by Spine Surgery Systematic Reviews: An AMSTAR-2 Evaluation From 4 Spine Journals

Joseph R. Dettori[1] , **Andrea C. Skelly[2], and Erika D. Brodt[2]**

Abstract

Study Design: A systematic cross-sectional survey of systematic reviews (SRs).

Objective: To evaluate the methodological quality of spine surgery SRs published in 2018 using the updated AMSTAR 2 critical appraisal instrument.

Methods: We identified the PubMed indexed journals devoted to spine surgery research in 2018. All SRs of spine surgical interventions from those journals were critically appraised for quality independently by 2 reviewers using the AMSTAR 2 instrument. We calculated the percentage of SRs achieving a positive response for each AMSTAR 2 domain item and assessed the levels of confidence in the results of each SR.

Results: We identified 28 SRs from 4 journals that met our criteria for inclusion. Only 49.5% of the AMSTAR 2 domain items satisfied the AMSTAR 2 criteria. Critical domain items were satisfied less often (39.1%) compared with noncritical domain items (57.3%). Domain items most poorly reported include accounting for individual study risk of bias when interpreting results (14%), list and justification of excluded articles (18%), and an a priori establishment of methods prior to the review or registered protocol (18%). The overall confidence in the results was rated "low" in 2 SRs and "critically low" in 26.

Conclusions: The credibility of a SR and its value to clinicians and policy makers are dependent on its methodological quality. This appraisal found significant methodological limitations in several critical domains, such that the confidence in the findings of these reviews is "critically low."

Keywords

systematic review, meta-analysis, AMSTAR, AMSTAR 2, spine surgery, intervention, research methodology

Introduction

The number of published systematic reviews (SRs) of studies of spinal interventions has increased rapidly over the past 20 years.[1] Clinical decision and policy makers rely on the results of high-quality SRs because they are generally recognized as representing the highest level of evidence. SRs are classified as observational research studies. As such, they are subject to a range of biases. The quality of a SR is dependent on how rigorously the research team attempts to limit these biases by following methodologically sound practices. It is important that users be able to distinguish high-quality reviews.

Several instruments have been designed to assess the quality of SRs based on established methodological processes for such reviews.[2-7] One popular instrument, the AMSTAR

(A MeaSurement Tool to Assess systematic Reviews), was first developed and published in 2007 to evaluate SRs of randomized controlled trials.[6] It was revised and updated in 2017 to include SRs of nonrandomized studies and to respond to critiques of the original. The result was the AMSTAR 2, a 16-item tool with a comprehensive user guide that provides an overall rating based on weaknesses in 7 critical domains.[8]

[1] Spectrum Research, Inc, Steilacoom, WA, USA
[2] Aggregate Analytics, Inc, Fircrest, WA, USA

Corresponding Author:
Joseph R. Dettori, Spectrum Research, Inc, Steilacoom, WA, USA.
Email: joe@specri.com

Furthermore, the AMSTAR 2 provides a scheme for interpreting weaknesses in both critical and noncritical domains and provides an overall confidence level in the results of the review.

Whereas we identified 1 study that assessed spine surgery meta-analyses using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist,[1] we have not identified any study that critically appraises SRs in the spine surgery literature using the AMSTAR 2. Therefore, the purpose of the current study is to evaluate the methodological quality of spine surgery SRs published in 2018 using the AMSTAR 2 critical appraisal instrument.

Methods

Study Selection

We identified PubMed indexed journals from the Web of Science's Journal Citation Reports in 2018 with an impact factor ≥2 and devoted to spine surgery research using the words "spine" or "spinal" in the title of the journal. The impact factor measures how often a publication is cited, not the quality of research. Also, we were interested in those journals with the highest citation frequency. We chose the year 2018 because it represented the latest complete year available at the beginning of our review. Each journal was searched for publications listed as a SR or meta-analysis. SRs were then screened to determine if they met a priori inclusion criteria. We included only those SRs that studied spine surgical interventions or adjunctive treatment to spine surgical interventions such as venous thromboembolism prophylaxis in spine surgery. We excluded reviews of prognostic factors, diagnostic criteria, therapy in nonoperative patients, and therapy in postoperative patients (eg, bracing or rehabilitation after spine surgery).

Study Evaluation

The quality of each SR was evaluated independently by 2 epidemiologists trained in SR methodology and critical appraisal (JD, AS) using the AMSTAR 2 (Table 1). Differences between evaluators in rating each domain item were identified. The evaluators then discussed the rationales for their respective score, and the differences were resolved when both evaluators agreed on a common assessment. The creators of AMSTAR 2 proposed 3 answer options for each item: "yes," "partial yes," or "no." For our purposes, we dichotomized the results, so that either a "yes" or "partial yes" denoted a positive result and "no" a negative result. Additionally, "no meta-analysis" was an option for items 11, 12, and 15. We then applied the AMSTAR 2 rating scheme for interpreting the overall confidence in the results of the SR (Table 2).

Data Analysis

The included SRs were summarized using descriptive statistics. We calculated the percentage of SRs achieving a positive response for each item and the percentage achieving each of

Table 1. 16-Item AMSTAR 2 Instrument: Critical Domains (Bold) Include Items 2, 4, 7, 9, 11, 13, and 15.

Item No.	Domain Questions
1	Did the research questions and inclusion criteria for the review include the components of PICO?
2	**Did the report of the review contain an explicit statement that the review methods were established prior to conduct of the review and did the report justify any significant deviations from the protocol?**
3	Did the review authors explain their selection of the study designs for inclusion in the review?
4	**Did the review authors use a comprehensive literature search strategy?**
5	Did the review authors perform study selection in duplicate?
6	Did the review authors perform data extraction in duplicate?
7	**Did the review authors provide a list of excluded studies and justify the exclusions?**
8	Did the review authors describe the included studies in adequate detail?
9	**Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?**
10	Did the review authors report on the sources of funding for the studies included in the review?
11	**If meta-analysis was justified did the review authors use appropriate methods for statistical combination of results?**
12	If meta-analysis was performed did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?
13	**Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?**
14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?
15	**If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?**
16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?

4 different levels of confidence in the results. For SRs without a meta-analysis, items 11, 12, and 15 did not contribute to the percentage calculations. In addition to providing overall results for all SRs, we stratified results by journal. Excel (Microsoft) was used for all analyses.

Results

Study Selection

Eight journals were listed in the Web of Science's 2018 Journal Citation Reports with "spine" or "spinal" in the title. We

Table 2. Rating Overall Confidence in the Results of the Review.

High: No or 1 noncritical weakness
The systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest
Moderate: More than 1 noncritical weakness[a]
The systematic review has more than 1 weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review
Low: One critical flaw with or without noncritical weaknesses
The review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest
Critically low: More than 1 critical flaw with or without noncritical weaknesses
The review has more than 1 critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies

[a] Multiple noncritical weaknesses may diminish confidence in the review, and it may be appropriate to move the overall appraisal down from moderate to low confidence.

excluded one for having published no SRs in 2018 (*Joint Bone Spine*) and 3 for having an impact factor less than 2.0 (*Spinal Cord*, *Clinical Spine Surgery*, and *The Journal of Spinal Cord Medicine*), leaving 4 journals meeting our inclusion criteria (Figure 1). A total of 72 studies were identified as SRs published in the 4 journals in 2018. A total of 44 SRs were excluded from this review; excluded studies and reasons for exclusion are listed in Table S1 in the supplemental material. The remaining 28 SRs met our criteria for inclusion: 13 from the *European Spine Journal*, 1 from the *Journal of Neurosurgery Spine*, 9 from *Spine*, and 5 from the *Spine Journal*.

Critical Appraisal of Systematic Reviews

The mean percentage of the AMSTAR 2 domain items satisfying the AMSTAR 2 criteria for all SRs was 49.5% (Figure 2). The critical domain items were satisfied less often (39.1%) compared with the noncritical domain items (57.3%). The SRs published in *The Spine Journal* had a higher percentage of both
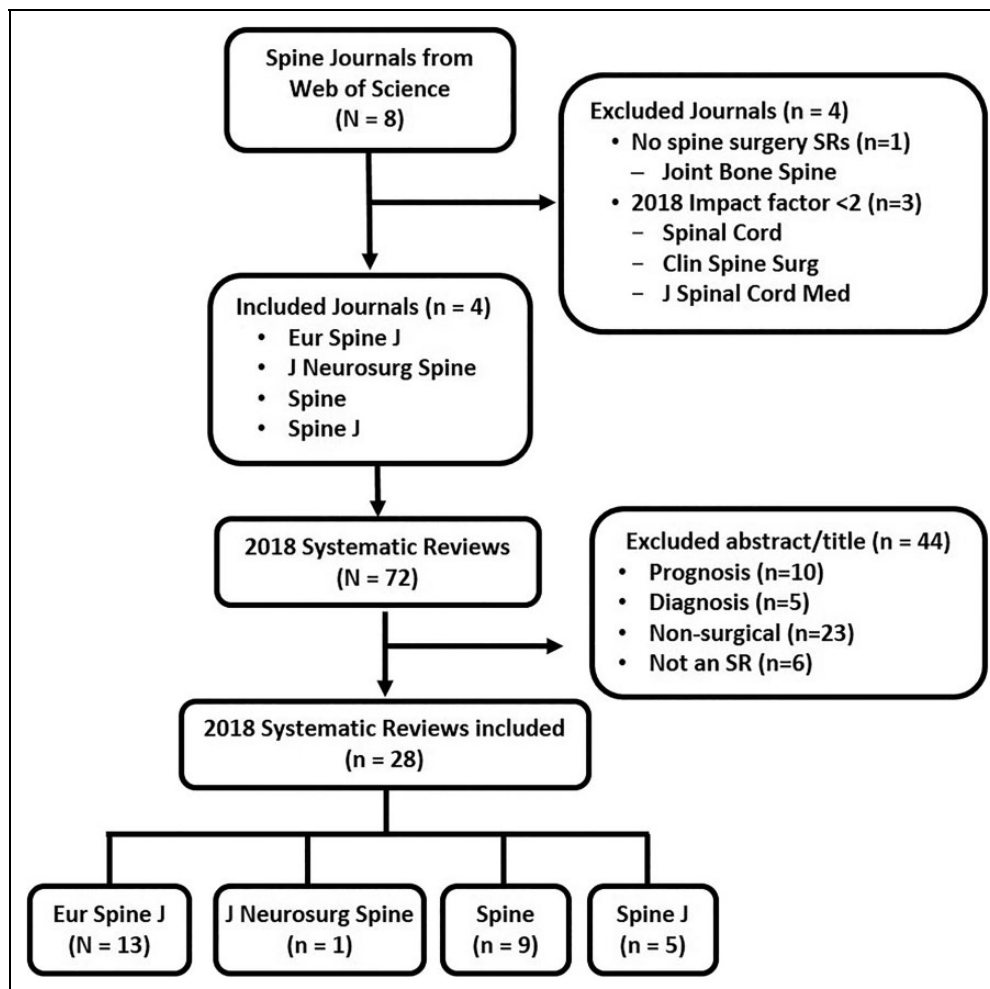

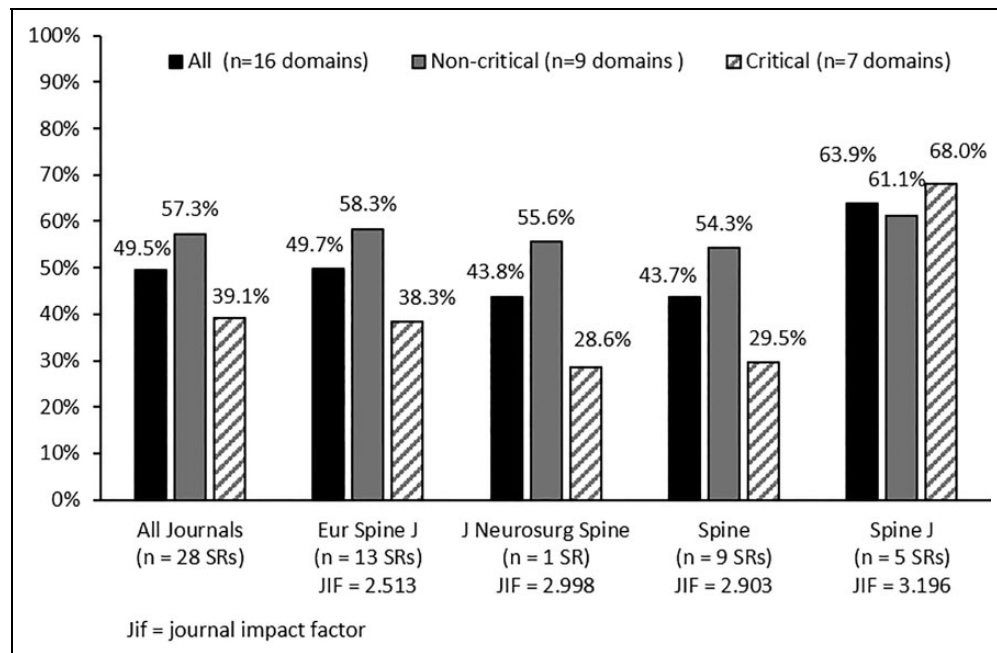**Figure 1.** Study selection. Abbreviation: SR, systematic review.

Figure 2. Mean percentage of AMSTAR 2 domain items satisfying the AMSTAR 2 criteria by journal. Abbreviations: JIF, journal impact factor; SR, systematic review.
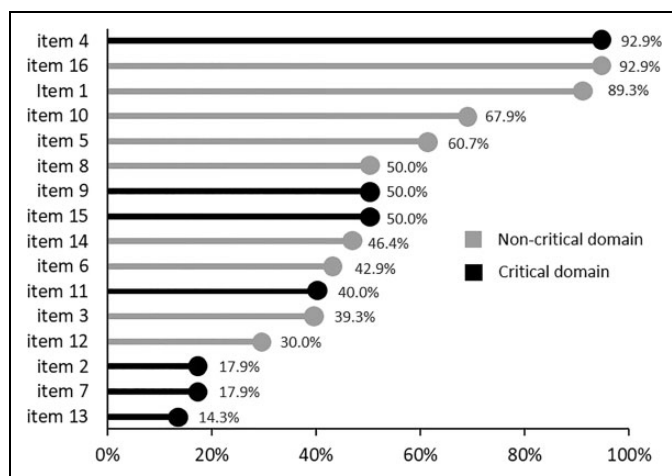

Figure 3. Percentage of systematic reviews satisfying the AMSTAR 2 criteria by domain item number.

critical and noncritical domain items answered satisfactorily compared with the other journals.

Of the 16 AMSTAR 2 domain items, 11 were reported ≤50% of the time across studies (Figure 3). Domain items best reported included comprehensive literature search (item 4, 93%), potential conflict of interest (item 16, 93%), and specification of inclusion criteria containing components of PICO (patient, intervention, comparator, and outcome; item 1, 89%). Domain items most poorly reported include accounting for individual study risk of bias (RoB) when interpreting the results (item 13, 14%), list and justification of excluded articles (item 7, 18%), and an a priori establishment of methods prior to the review or registered protocol (item 2, 18%).

Table 3. Confidence in the Results of the Systematic Review.

	High, n (%)	Moderate, n (%)	Low, n (%)	Critically Low, n (%)
Overall	0 (0.0)	0 (0.0)	2 (7.1)	26 (92.9)
European Spine Journal	0 (0.0)	0 (0.0)	1 (7.7)	12 (92.3)
Journal of Neurosurgery Spine	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)
Spine	0 (0.0)	0 (0.0)	0 (0.0)	9 (100.0)
Spine Journal	0 (0.0)	0 (0.0)	1 (20.0)	4 (80.0)

Overall Confidence in Results of Systematic Reviews

The overall confidence in the results of the 28 reviews was rated as "low" in 2 SRs[9,10] and "critically low" in the rest[11-36] (Table 3). Critical appraisal of each SR is found in Table S2 in the supplemental material.

Discussion

The included 28 SRs published in 4 high-impact spine journals in 2018 demonstrated poor compliance with the AMSTAR 2 instrument; confidence in the results was graded as "critically low" in 93% (n = 26). To our knowledge, this is the first critical appraisal of spine surgery SRs.

A few AMSTAR 2 items were well reported among the included sample of SRs. Almost all SRs used a comprehensive electronic literature search to identify potential articles. It is our experience that many spine surgeons equate SRs with systematic literature searching, underestimating the importance of the other aspects of a SR. This could, in part, account for the high percentage correctly reporting this

domain item while at the same time neglecting many other aspects of the SR that document sound methodological principles. Almost all review authors declared whether they had potential conflict of interest. We expected a high percentage with this domain item because it is required now by many journals. However, although there has been an emphasis on the importance of declaring financial conflicts of interest, little attention appears to be given to declaring nonfinancial conflicts of interest (NFCOIs) such as personal relationships, strongly held beliefs, and the desire for career advancement. When ignored, NFCOIs can call into question the impartiality of a review.[37]

Whereas nearly 70% of the AMSTAR 2 domain items were reported ≤50% of the time, 3 items were reported <18% of the time: publishing a protocol prior to initiating the review, detailed identification of excluded articles, and accounting for RoB in the results and conclusions. Publishing of a protocol prior to initiating a SR is often not considered by those performing spine surgery SRs. The methods for conducting the review, including inclusion/exclusion criteria and planned analyses, should be developed a priori and stated as such at a minimum. A SR is an observational study and, therefore, requires agreed upon methods prior to starting the review. Adhering to the methods helps reduce RoB.

Listing of excluded articles and the justification for exclusion were also poorly reported. Several articles listed general reasons for exclusion in their PRISMA flow diagram but failed to identify the specific studies excluded. Without explicit citation of the articles excluded, the potential impact of their exclusion cannot be known.

Half of the reviews used appropriate techniques in assessing the RoB in primary studies. However, only 14% accounted for the RoB when interpreting the results or drawing conclusions. Many SRs drew conclusions from the data of studies with a high RoB and did not account for the poor quality of the data available from such studies. As a result, conclusions drawn were not supported by the quality of the data. The spine surgery literature often includes studies with varying levels of quality. When presenting the data in a SR, investigators should emphasize the highest-quality studies, those with the lowest RoB, and acknowledge that the confidence in the data is likely low when high-RoB studies are included.

Previous assessments of SR quality using different critical appraisal instruments across a variety of surgical subspecialties have demonstrated similar shortcomings as the SRs we reviewed.[38-40] For example, one study of the neurosurgery literature[39] reported only that 18% of SRs listed the excluded studies, and just 21% assessed publication bias using the original AMSTAR instrument. In our review, the journal with the highest impact factor had the highest percentage of AMSTAR 2 domain items satisfying the AMSTAR 2 criteria. This is consistent with other reviews assessing quality by journal impact factor.[1,41]

Our analysis is limited to SRs of spine surgery published in 2018 in 4 spine journals with an impact factor >2.0. Our review did not evaluate aspects of SR quality outside of the domains of the AMSTAR 2 instrument. By dichotomizing whether criteria were met or partially met as "yes" versus not met, the overall quality of some reviews may have been rated higher than if a yes, partially met, or no scale had been used. We did not stratify studies by type of surgery or focus (eg, complications vs effectiveness); however, the methodological rigor (and appraisal) for SRs should not differ.

Conclusions

As with any research, the credibility of a SR and its value to clinicians and policy makers are dependent on its methodological quality. Although the SRs included came from spine journals with higher impact factors, this appraisal found that confidence in the findings of these reviews was critically low.

References

1. Evaniew N, van der Watt L, Bhandari M, et al. Strategies to improve the credibility of meta-analyses in spine surgery: a systematic survey. *Spine J*. 2015;15:2066-2076.
2. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc*. 2015;13:132-140.
3. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature: VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA*. 1994;272:1367-1371.
4. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*. 1991;44:1271-1278.
5. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316:450-455.
6. 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.
7. Whiting P, Savovic J, Higgins JP, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-234.

8. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008.

9. Gao S, Lv Z, Fang H. Robot-assisted and conventional free-hand pedicle screw placement: a systematic review and meta-analysis of randomized controlled trials. *Eur Spine J*. 2018;27: 921-930.

10. Sarode DP, Demetriades AK. Surgical versus nonsurgical management for type II odontoid fractures in the elderly population: a systematic review. *Spine J*. 2018;18:1921-1933.

11. Dijkerman ML, Overdevest GM, Moojen WA, Vleggeert-Lankamp CLA. Decompression with or without concomitant fusion in lumbar stenosis due to degenerative spondylolisthesis: a systematic review. *Eur Spine J*. 2018;27:1629-1643.

12. Guan J, Chen Z, Wu H, et al. Is anterior release and cervical traction necessary for the treatment of irreducible atlantoaxial dislocation? A systematic review and meta-analysis. *Eur Spine J*. 2018;27:1234-1248.

13. Jing Z, Dong J, Li Z, Nan F. Single balloon versus double balloon bipedicular kyphoplasty: a systematic review and meta-analysis. *Eur Spine J*. 2018;27:2550-2564.

14. Makanji H, Schoenfeld AJ, Bhalla A, Bono CM. Critical analysis of trends in lumbar fusion for degenerative disorders revisited: influence of technique on fusion rate and clinical outcomes. *Eur Spine J*. 2018;27:1868-1876.

15. Malik AT, Jain N, Kim J, Khan SN, Yu E. Sexual activity after spine surgery: a systematic review. *Eur Spine J*. 2018;27: 2395-2426.

16. Malik AT, Panni UY, Mirza MU, Tetlay M, Noordin S. The impact of surgeon volume on patient outcome in spine surgery: a systematic review. *Eur Spine J*. 2018;27:530-542.

17. Morris MT, Tarpada SP, Cho W. Bone graft materials for posterolateral fusion made simple: a systematic review. *Eur Spine J*. 2018;27:1856-1867.

18. Qin R, Chen X, Zhou P, Li M, Hao J, Zhang F. Anterior cervical corpectomy and fusion versus posterior laminoplasty for the treatment of oppressive myelopathy owing to cervical ossification of posterior longitudinal ligament: a meta-analysis. *Eur Spine J*. 2018;27:1375-1387.

19. Shao ZX, Fang X, Lv QB, et al. Comparison of combined anterior-posterior approach versus posterior-only approach in neuromuscular scoliosis: a systematic review and meta-analysis. *Eur Spine J*. 2018;27:2213-2222.

20. Tan JH, Liu G, Ng R, Kumar N, Wong HK, Liu G. Is MIS-TLIF superior to open TLIF in obese patients? A systematic review and meta-analysis. *Eur Spine J*. 2018;27:1877-1886.

21. Thakar C, Kieser DC, Mardare M, Haleem S, Fairbank J, Nnadi C. Systematic review of the complications associated with magnetically controlled growing rods for the treatment of early onset scoliosis. *Eur Spine J*. 2018;27:2062-2071.

22. Zanirato A, Damilano M, Formica M, et al. Complications in adult spine deformity surgery: a systematic review of the recent literature with reporting of aggregated incidences. *Eur Spine J*. 2018;27:2272-2284.

23. Alvi MA, Alkhataybeh R, Wahood W, et al. The impact of adding posterior instrumentation to transpsoas lateral fusion: a systematic review and meta-analysis. *J Neurosurg Spine*. 2018;30:211-221.

24. Ahmed AT, Abdel-Rahman O, Morsy M, et al. Management of sacrococcygeal chordoma: a systematic review and meta-analysis of observational studies. *Spine (Phila Pa 1976)*. 2018;43: E1157-E1169.

25. Buerba RA, Sharma A, Ziino C, Arzeno A, Ajiboye RM. Bisphosphonate and teriparatide use in thoracolumbar spinal fusion: a systematic review and meta-analysis of comparative studies. *Spine (Phila Pa 1976)*. 2018;43:E1014-E1023.

26. Daniel JW, Botelho RV, Milano JB, et al. Intraoperative neurophysiological monitoring in spine surgery: a systematic review and meta-analysis. *Spine (Phila Pa 1976)*. 2018;43: 1154-1160.

27. Ghasem A, Sharma A, Greif DN, Alam M, Maaieh MA. The arrival of robotics in spine surgery: a review of the literature. *Spine (Phila Pa 1976)*. 2018;43:1670-1677.

28. Lu VM, Ho YT, Nambiar M, Mobbs RJ, Phan K. The perioperative efficacy and safety of antifibrinolytics in adult spinal fusion surgery: a systematic review and meta-analysis. *Spine (Phila Pa 1976)*. 2018;43:E949-E958.

29. Mosenthal WP, Landy DC, Boyajian HH, et al. Thromboprophylaxis in spinal surgery. *Spine (Phila Pa 1976)*. 2018;43: E474-E481.

30. Oliver JD, Goncalves S, Kerezoudis P, et al. Comparison of outcomes for anterior cervical discectomy and fusion with and without anterior plate fixation: a systematic review and meta-analysis. *Spine (Phila Pa 1976)*. 2018;43:E413-E422.

31. Osman NS, Cheung ZB, Hussain AK, et al. Outcomes and complications following laminectomy alone for thoracic myelopathy due to ossified ligamentum flavum: a systematic review and meta-analysis. *Spine (Phila Pa 1976)*. 2018;43:E842-E848.

32. Yoshihara H. Surgical treatment of Lenke type 5 adolescent idiopathic scoliosis: a systematic review. *Spine (Phila Pa 1976)*. 2019;44:E788-E799.

33. Huysmans E, Goudman L, Van Belleghem G, et al. Return to work following surgery for lumbar radiculopathy: a systematic review. *Spine J*. 2018;18:1694-1714.

34. Levin JM, Tanenbaum JE, Steinmetz MP, Mroz TE, Overley SC. Posterolateral fusion (PLF) versus transforaminal lumbar interbody fusion (TLIF) for spondylolisthesis: a systematic review and meta-analysis. *Spine J*. 2018;18:1088-1098.

35. Noorian S, Sorensen K, Cho W. A systematic review of clinical outcomes in surgical treatment of adult isthmic spondylolisthesis. *Spine J*. 2018;18:1441-1454.

36. Yang X, Janssen T, Arts MP, Peul WC, Vleggeert-Lankamp CLA. Radiological follow-up after implanting cervical disc prosthesis in anterior discectomy: a systematic review. *Spine J*. 2018;18: 1678-1693.

37. Agency for Healthcare Research and Quality. Identifying and managing nonfinancial conflicts of interest for systematic reviews. https://effectivehealthcare.ahrq.gov/products/conflict-management/research. Published May 23, 2013. Accessed March 23, 2020.

38. Dijkman BG, Abouali JA, Kooistra BW, et al. Twenty years of meta-analyses in orthopaedic surgery: has quality kept up with quantity? *J Bone Joint Surg Am*. 2010;92:48-57.

39. Klimo P Jr, Thompson CJ, Ragel BT, Boop FA. Methodology and reporting of meta-analyses in the neurosurgical literature. *J Neurosurg*. 2014;120:796-810.

40. Sharma R, Vannabouathong C, Bains S, et al. Meta-analyses in joint arthroplasty: a review of quantity, quality, and impact. *J Bone Joint Surg Am*. 2011;93:2304-2309.

41. Fleming PS, Koletsi D, Seehra J, Pandis N. Systematic reviews published in higher impact clinical journals were of higher quality. *J Clin Epidemiol*. 2014;67:754-759.