Review article

Quality of meta-analyses in major leading orthopedics journals: A systematic review

X. Zhi\textsuperscript{a,1}, Z. Zhang\textsuperscript{a,1}, J. Cui\textsuperscript{b}, X. Zhai\textsuperscript{b}, X. Chen\textsuperscript{b,*, J. Su}\textsuperscript{b,*}

\textsuperscript{a} Graduate Management Unit, Shanghai hospital affiliated to the Second Military Medical University, Shanghai road, 200433 Shanghai, PR, China
\textsuperscript{b} Department of Orthopedics, Shanghai hospital affiliated to the Second Military Medical University, Shanghai road, 200433 Shanghai, PR, China

\section*{ARTICLE INFO}

\textbf{Article history:}
Received 28 February 2017
Accepted 24 August 2017

\textbf{Keywords:}
Meta-analysis
Methodological quality
Orthopedics
PRISMA
AMSTAR

\section*{ABSTRACT}

\textbf{Background:} Meta-Analyses are the basis of professional and healthcare agencies recommendations and have a growing importance. Quality of meta-analyses has been investigated in some medical fields but to our best knowledge this issue remains under investigated in orthopedics. Therefore, we performed a systematic analysis to: 1) after the introduction of PRISMA statement as a comprehensive guideline and the use of the AMSTAR tool as the standard for sufficient review methodology, has the quality of MAs improved because of that? 2) have some general characteristics influenced the quality of MAs (country, funding source, number of authors)?

\textbf{Material and Methods:} We systematically searched the meta-analyses in the top four journals with the impact factor (2015) as following: JBJ, Osteoarthritis Cartilage Arthroscopy and Clin Orthop Relat Res from 2005 to 2008 and from 2012 to 2015. Likewise from 2012–2015, we also analyzed the meta-analyses from OTSR, Characteristics were extracted based on the PRISMA statement and the AMSTAR tool. Country, number of authors, funding source were also extracted.

\textbf{Results:} A total of 154 meta-analyses were included in the present study. Score with PRISMA statement and the AMSTAR checklist were 20.86 ± 3.04 out of a maximum of 27 and 7.86 ± 1.55 out of a maximum of 11. The best journal was OTSR according to the PRISMA (23.06 ± 1.92) and AMSTAR (9.13 ± 0.87) scores. And the worst journal was Clin Orthop Relat Res according to the PRISMA score (19.4 ± 2.70) and JBJS according to the AMSTAR score (6.78 ± 1.65). Twelve items showed significant difference in the PRISMA statement, and five items in the AMSTAR checklist. Integral score of PRISMA statement and AMSTAR checklist has a significant difference between 2005–2008 and 2012–2015. The MAs reported from U.S. (56, 36.4%) were more than any other region in the world. And the MAs published by Asia/Oceania increased remarkably between these two period times [from (4, 10.8%) to (45, 38.5%)].

\textbf{Conclusion:} This study showed that methodological reporting quality of meta-analyses in the major orthopedics journals has improved after the publication of the PRISMA statement.

\textbf{Level of evidence:} Level III.

\textsuperscript{©} 2017 Elsevier Masson SAS. All rights reserved.

\textbf{Abbreviations:} MA, Meta-analysis; SRs, System reviews; PRISMA, Preferred reporting items for systematic reviews and meta-analyses; AMSTAR, Assessing the methodological quality of systematic review; JBJS, The Journal of Bone and Joint Surgery; Osteoarthritis Cartilage, Osteoarthritis and Cartilage; Arthroscopy, The journal of arthroscopic and related surgery; CORR, Clinical Orthopedics and Related Research; OTSR, Orthopedics, Traumatology & Surgery Research.

\textsuperscript{1} Corresponding authors. Department of Orthopedics, Shanghai hospital affiliated to the Second Military Medical University, Shanghai, China.

\textbf{E-mail addresses:} sirchenxiao@126.com (X. Chen), dsrujicani@163.com (J. Su).

\textsuperscript{*} These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.otsr.2017.08.009
1877-0568(17)30168-3

1. Introduction

Systematic reviews (SRs) and meta-analyses (MAs) have become important in health care. Clinicians refer to them to learn the latest clinical methods in these fields [1]. MAs and SRs are indispensable tools for summing up evidence precisely and reliably. MAs and SRs help clinicians keep up to date; supply evidence for policy makers to estimate risks, interests and harms of healthcare behaviors and treatments; collect and sum up related researches for
patients; provide a beginning point for clinical practice guideline
developers; summarize of previous researches for funders wishing
to support new studies; and help editors select the merits of pub-
lishing reports of new researches [2]. It was estimated that at least
2500 new MAs (including systematic reviews, SRs) were indexed
in Medline yearly [3]. However, lots of MAs were methodologically
poorly reported and could introduce bias and impair the reliability
of conclusions although rigorous methodology is a feature of MAs
[4,5].

To prevent such incidents, the PRISMA (Preferred Reporting
Items for Systematic Reviews and Meta-Analyses, PRISMA) was
developed in 2009 (Supplemental data 1) [1]. It was an evidence-
based minimum set of items for reporting in Systematic reviews
and meta-analyses assessing randomized control trials (RCT) [6].
The PRISMA statement was developed by a group of 29 review
authors, methodologists, clinicians, medical editors, and con-
sumers [1]. The objective of the PRISMA statement was to help
authors improve the reporting quality of SRs and MAs. PRISMA
could also be advantageous for crucial appraisal of published sys-
tematic reviews. It has been extensively used as a criterion to assess
the reporting quality of MAs since it was published in 2009 [1].
AMSTAR (Assessing the Methodological Quality of SR, AMSTAR) also
emerged in 2007 (Supplemental data 2) [7], and subsequently was
simplified to 11 items from 37 items. The aim of the AMSTAR is
to evaluate the reporting quality of SRs, by establish upon empirical
data on previously reported tools, empirical evidence and utilizing
expert opinion [8]. After the PRISMA statement and AMSTAR tool
were reported, assessment of the reported quality of MAs have been
reported in many fields, including surgery [9,10], pulmonary dis-
ease [11], telerehabilitation [12], nursing [13] and gastroenterology
and hepatology [6].

Meta-Analysis with its present form of statistically integrating
information from several studies all with a common underlying
theme has been around for over 25 years [14]. There are growing
numbers of MAs being published in orthopedics fields, and guiding
the clinical work. However, to our best knowledge, the evaluation of
orthopedic journals has never been reported before. To solve this
issue, we systematically evaluated the methodological reporting
quality of MAs before and after the publication of the developed
PRISMA statement (from 2005 to 2008 and from 2012 to 2015), in
order to observe the quality of MAs whether improve or not after
the standard was published. MAs were selected from the 4 ortho-
edics journals with the highest impact factor (2015) and Orthopaedics & Traumatology: Surgery & Research which is not
a leading journal in this field but also introduced PRISMA before
2012.

We introduce the following questions:

• has the quality of MAs improved after the introduction of
PRISMA statement as a comprehensive guideline and the use
of the AMSTAR tool as the standard for sufficient review
methodology?
• have some general characteristics influenced the quality of MAs
(country, funding source, number of authors)?

2. Methods of search strategy and criteria

2.1. Search strategy

To appraise the methodological quality of MAs on orthopedics
fields, we selected the top four journals with the highest impact
Osteoarthritis and Cartilage (Osteoarthritis Cartilage), Arthroscopy
– the journal of arthroscopic and related surgery (Arthroscopy) and
Clinical Orthopaedics and Related Research (CORR) [15]. In

addition, we picked up the MAs published from 2012 to 2015 in
Orthopaedics & Traumatology: Surgery & Research (OTSR), another
journal, which has also introduced PRISMA before 2012 as com-
parison.

The conformed MAs were conducted on the PubMed/MEDLINE
database [16] by two co-first authors on October 1st, 2016. The
ISSNs (Print) of five journals were 1535–1386, 1063–4584,
0749–8063, 0009–921X and 1877–0568. To evaluate the PRISMA
statement published in 2009, we chose the articles from 2005 to
2008 and from 2012 to 2015. We selected the “Meta-analysis” in
the “Article types” project of Web of science, and excluded the fol-
lowing types: randomized controlled trial or review or the article

2.2. Data extraction and criteria

We primarily extracted geography, publishing, clinical and epide-
miological characteristics based on the previous meta-analysis and
systematic reviews, for example region/country, number of
authors, publishing year, number of patients and funding source.
Then we collected the information of MAs with PRISMA statement
[2] and AMSTAR checklist [7]. The PRISMA statement is a 27-item
list and the AMSTAR checklist is an 11–item list. The work of extrac-
tion information was completed by two co-first authors, and any
disagreement was resolved by discussion between the two review-
ers or other authors.

2.3. Statistical analyses

All statistical analyses were performed with SPSS 21.0 (SPSS Sta-
tistical software version 21.0. IBM Corporation, USA). Frequencies
and proportions were analyzed, and descriptive statistics (mean,
standard deviation) of the country, funding source were employed
(Table 1). For the parabole between years and journals, we per-
formed chi-squared test and fisher exact test (Tables 2 and 3).
For the determination of variables that associated with PRISMA
or AMSTAR score, we firstly accomplished the univariate linear
regression and then combined with stepwise backward multiple
regression analysis. P < 0.05 was regarded statistically significant.
Effect size (Odds ratio, OR) and 95% confidence intervals were
reported for significant associations (Table 4).

3. Results

3.1. Search results

A total of 167 MAs were retrieved from the five journals, in which
13 studies were excluded (Fig. 1) [all detailed results of the 167

Table 1

<table>
<thead>
<tr>
<th>Principal characteristics of the included trials.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Number</td>
</tr>
<tr>
<td>No. authors</td>
</tr>
<tr>
<td>1–3</td>
</tr>
<tr>
<td>4–6</td>
</tr>
<tr>
<td>&gt; 6</td>
</tr>
<tr>
<td>Region/Country</td>
</tr>
<tr>
<td>US</td>
</tr>
<tr>
<td>Europe</td>
</tr>
<tr>
<td>Asia/Oceania</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>Funding source</td>
</tr>
<tr>
<td>Industry</td>
</tr>
<tr>
<td>Public</td>
</tr>
<tr>
<td>Not specific</td>
</tr>
<tr>
<td>None</td>
</tr>
</tbody>
</table>
studies investigated are reported in Supplemental data 3). Finally, 154 MAs were accepted in the article, including 32 from the JBJS, 34 from the Arthroscopy, 41 from the CORR, 31 from the Osteoarthritis Cartilage and 16 from the OTSR. Thirty-seven (24.0%) MAs were published between 2005 and 2008, and one hundred and seven (76.0%) MAs were published between 2012 and 2015 (all detailed results of the 154 inclusions are reported in Supplemental data 3).

### 3.2. General characteristics

Crucial characteristics of the included MAs were shown in the Table 1.87 (56.5%) of MAs have 4-6 authors per article. As a whole, the MAs reported from US (56, 36.4%) were more than any other region in the world. The Asia/Oceania production increased remarkably between these two periods [from 4, 10.8% to 45, 38.5%). In 2005–2008, 18.9% of the MAs were with funding source, and 32.4% were not specific. And it was 42.7% with the funding source and 16.2% that were not specific in 2012–2015.

#### 3.3. Compliance with the PRISMA checklist

The proportion and percentage of MAs that adequately reported each PRISMA items were showed in Fig. 2 and Table 2. Summary of evidence (item 24) was reported 100%. Most domains were
Table 4: Univariate and multivariate exploratory linear relationship of changes in PRISMA score and AMSTAR score.

<table>
<thead>
<tr>
<th>Items</th>
<th>PRISMA Score</th>
<th>AMSTAR Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate exploratory variable</td>
<td>sβ (OR)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe (vs. America)</td>
<td>0.54</td>
<td>(-0.780, 1.86)</td>
</tr>
<tr>
<td>Asia/Oceania (vs. Europe)</td>
<td>1.676</td>
<td>(0.523, 2.829)</td>
</tr>
<tr>
<td>Others</td>
<td>1.466</td>
<td>(-0.131, 3.064)</td>
</tr>
<tr>
<td>Journal (vs. Osteoarthritis Cartilage)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthroscopy (vs. JBJS)</td>
<td>0.433</td>
<td>(-1.001, 1.866)</td>
</tr>
<tr>
<td>CORR (vs. JBJS)</td>
<td>-0.156</td>
<td>(-1.523, 1.210)</td>
</tr>
<tr>
<td>OTSR (vs. JBJS)</td>
<td>0.769</td>
<td>(-0.643, 2.181)</td>
</tr>
<tr>
<td>Funding (vs. Industry)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public (vs. None)</td>
<td>0.500</td>
<td>(-1.204, 2.204)</td>
</tr>
<tr>
<td>None (vs. Public)</td>
<td>0.273</td>
<td>(-1.493, 2.040)</td>
</tr>
<tr>
<td>Numbers of authors (vs. 1–3)</td>
<td>0.143</td>
<td>(-0.091, 0.378)</td>
</tr>
<tr>
<td>Year (vs. 2005–2008)</td>
<td>2.807</td>
<td>(1.761, 3.853)</td>
</tr>
<tr>
<td>Multivariate exploratory variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year (vs. 2005–2008)</td>
<td>14.750</td>
<td>(12.107, 17.393)</td>
</tr>
</tbody>
</table>

*P<0.05.

Fig. 1. Flow of information through the different phases of a systematic review. (JBJS = the Journal of Bone and Joint Surgery; Osteoarthritis Cartilage = Osteoarthritis and Cartilage; Arthroscopy = The Journal of Arthroscopic and Related Surgery; CORR = Clinical Orthopaedics and Related Research; OTSR = Orthopaedics & Traumatology: Surgery & Research).

Fig. 2. The compliance with the PRISMA Checklist in 2005–2008 and 2012–2015. There are 12 items significantly improvements in 2012–2015 than those in 2005–2008. The PRISMA score was 21.50 ± 2.69 in 2012–2015 and 18.27 ± 3.03 in 2005–2008 (P = 0.05).

However, there were some items poorly reported, including risk of bias in individual studies (item 12, 55, 35.7%), risk of bias across studies (item 15, 42, 27.3%), additional analyses (item 16, 35, 22.7%). Risk of bias within studies (item 19, 56, 36.4%), risk of bias across studies (item 22, 45, 29.2%) and Additional analyses (item 23, 38, 24.7%).

There were significant improvements in 2012–2015 than those in 2005–2008. It was reflected in the following items: Structured summary (item 2, P = 0.003), Protocol and registration (item 5, P = 0.03), Eligibility criteria (item 6, P = 0.001), Information sources (item 7, P < 0.001), Search (item 8, P = 0.002), Data collection process (item 10, P = 0.041), Data items (item 11, P = 0.001), Summary measures (item 13, P = 0.001), Additional analyses (item 16, P = 0.007), Study characteristics (item 18, P < 0.001), Results of individual studies (item 20, P = 0.004) and Synthesis of results (item 21, P = 0.013). The PRISMA score was 21.50 ± 2.69 in 2012–2015 (77% of items adequately reported, on average), has significant improvements compared with the PRISMA score in 2005–2008 (18.27 ± 3.03, P < 0.001). In addition, OTSR performed the best in domains of the average score of PRISMA (23.06 ± 1.92) and risk of bias in individual studies (item 12, 15, 93.8%). JBJS performed the best in domains of Funding (item 27, 31, 96.9%).

3.4. Compliance with the AMSTAR Checklist

Fig. 3 and Table 3 showed the compliance of the reporting items with the AMSTAR checklist. The “a priori” design (item 1, 154, 100%) was reported best. Duplicate study selection and data extraction (item 2, 149, 96.8%) and Characteristics of studies (item 6, 139, 90.3%) were reported well. However, Grey literature search (item 4: 28, 18.2%) and Publication bias assessed (item 10, 49, 31.8%) were poorly reported.

Significant improvements were found in 2012–2015 compared with it in 2005–2008 in some domains, including Duplicate study selection and data extraction (item 2, P = 0.012), List of studies (included and excluded) (item 5, P < 0.001), Characteristics of studies (item 6, P < 0.001), Scientific quality in conclusions (item 8, P = 0.004) and Methods to combine the studies (item 9, P = 0.002). In addition, Osteoarthritis Cartilage reported best of scientific quality in conclusions (item 8, P < 0.001) and OTSR reported best of Comprehensive literature search (item 3, P = 0.008), Scientific quality assessed (item 7, P < 0.001) and conflict of interest (item 11, P < 0.001). OTSR performed the best in domains of comprehensive...
Literature search (item 3, 16, 100%), conflict of interest (item 11, 16, 100%) and performed the worst in Grey literature search (item 4, 0.0%). And arthroscopy performed the best in domains of scientific quality in conclusions (item 8, 34, 100%) and performed the worst in conflict of interest (item 11, 0, 0%). The AMSTAR score was 8.20 $\pm$ 1.38 in 2012–2015 (74% of items adequately reported, on average), which was improved when compared with the AMSTAR score in 2005–2008 (6.54 $\pm$ 1.39, $P < 0.001$).

3.5. Variables associated with PRISMA and AMSTAR reporting

Results of univariate exploratory linear regression and multivariate exploratory variable regression were presented in Table 4. Univariate exploratory linear regression showed that publication of articles from the Asia/Oceania increased in 1.676 score of PRISMA score compared with that from the America (95% CI: 0.523–2.829, $P = 0.005$) and increased in 1.097 score of AMSTAR score compared with that from the America (95% CI: 0.526–1.668, $P < 0.001$). And publication of articles supported by public support in increased in 0.421 score of AMSTAR score compared with that supported by the industry (95% CI: $-0.430$–1.272, $P < 0.001$). Publication of articles from OTSR increased in 3.214 score (95% CI: 1.454–4.974, $P < 0.001$) of PRISMA score and 0.973 score (95% CI: 0.126–1.821, $P = 0.025$) of AMSTAR score compared with that from Osteoarthritis Cartilage. Multivariate exploratory variable regression showed that publication of articles between 2012 and 2015 were related with an increase in 14.750 score of PRISMA score compared with that in 2005–2008 (95% CI: 12.107–17.393, $P < 0.001$). Similarly, articles published in 2012–2015 had a related with an increase in 5.660 score of AMSTAR score compared with 2005–2008 (95% CI: 4.395–6.925, $P = 0.001$). Points of PRISMA scores and AMSTAR scores of each journal from different years were given in Fig. 4.

4. Discussion

To the best of our knowledge, this is the first study to describe the methodological reporting quality of MAs publications in five orthopedics journals from 2005 to 2008 and from 2012 to 2015, including JBJS, CORR, Arthroscopy, Osteoarthritis Cartilage and OTSR. 56.5% of the 154 MAs have been completed by 4–6 authors, and most articles published by the US accounting for 36.4%. The articles from the Asia/Oceania were increased from 10.8% in 2005–2008 to 38.5% in 2012–2015, in which China made a great contribution. Due to the support of a variety of funds from public and industries, increasing number of Chinese scientists have participated in this promising research field [17]. In 2015, the amount of research spent in China was 14,000 hundred million Yuan, 2.10% of the GDP, which was increased by 1148.2 hundred million Yuan in 2012 [18].

There are several limitations in the present study. First, there are numerous journals in the orthopedics field, but this study only selected the five journals to represent the quality of MAs of all journals, and there may be with a little error. Second, only MAs published in eight years were included. The study did not show tendency in the quality of these journals consecutively. Besides, there may be some high-quality MAs were missed. Third, in this study, the methodology was the only criteria, it may result as some well–designed and well–conducted MAs were missed. However, to some extent, this study reflected the reporting quality of MAs in the orthopedic field.

In most cases, PRISMA statements are important for authors to improve their article's quality and transparency [19]. Although there was huge promotion of quality of MAs published in 2012–2015 compared with that in 2005–2008, there were still many shortcomings in some items. Twelve items were significantly improved in 2012–2015 compared with that in 2005–2008. For example, Structured summary (item 2) increased from 75.7% to 94.0%, and Data items (item 11) increased from 56.8% to 83.8%. And there was significant improvement in PRISMA score, from 18.27 to 21.50 in average score. But the difference of PRISMA score in two period times is tiny. We think the MAs published in 2005–2008 also have a high quality and only have shortcomings in some items, and with the publication of PRISMA, in 2012–2015, the quality of MAs has comprehensive progress. It was shown that many authors of MAs paid much more attention to the methodological reporting quality of MAs. But there were also some items unsatisfactory, such as Risk of bias in individual studies (item 12), Risk of bias across studies (item 15), Risk of bias within studies (item 19), Risk of bias across studies (item 22) and Additional analyses (item 23), there was no obvious difference between 2012–2015 and 2005–2008 in these items. Although selection method approached to assess and adjusted for publication biases in meta-analysis have a long history that dates back over 30 years, they have yet to be subject to an extensive simulation study that covers multiple continuously related settings and that evaluated model performance across multiple dimensions, we should continue to improve the quality of items for bias [20]. Interestingly, the items that were associated with “Additional analyses” both in the methods (item 16) were still in lower level, despite the fact that there were significant differences from 2012–2015 to 2005–2008. Additional analyses include
sensitivity analysis, grouping analyses and meta-regression. It can help understand whether the results of their reviews were robust or dependable [21,22]. It represented that the PRISMA guidelines wasn’t completely used by the authors. So, we should advocate publicizing the PRISMA guidelines through forum, lectures and academic conference etc., to make the quality of MAs gradually increased.

Apart from PRISMA guidelines, AMSTAR was also an important and effective measurement tool to appraise the methodological quality of meta-analyses [23,24]. There were 4 items had a significant difference in AMSTAR score from 2012–2015 to 2005–2008. Among them List of studies included and excluded (item 5) has a high degree of promotion. Meanwhile, the AMSTAR score also has a significant improvement, explaining that AMSTAR played an important role in improving the quality of MAs. But, there were a few problems. First, Publication bias assessed (item 10) in 2012–2015 was better than before (24.3% in 2005–2008), but still at a low level (34.2% in 2012–2015). Second, Comprehensive literature search (item 3), Grey literature search (item 4), Publication bias assessed (item 10) and Conflict of interest (item 11) have no significant change between two period times. Including grey literature can broaden the scope to relevant studies, thereby providing a more complete view of available evidence, even though searching for grey literature can be challenging [25]. So, popularize of AMSTAR score was most important. Interestingly, the AMSTAR score of four journals has a statistical difference, and the score of OTSR was better than others.

Multivariate regression indicated that MAs published in 2012–2015 had a relation with an additional 2.807 in PRISMA score and 1.528 in AMSTAR score to those published in 2005–2008. We could see the growth trend of the points of PRISMA scores and AMSTAR scores of each journal in two period times. After the PRISMA score was published in 2009, it might take an important role in improving the methodological quality of MAs [9].

5. Conclusion

In summary, the present study showed that the methodological quality of MAs from five journals. Although, it has shown a great trend to improve in some items in PRISMA and AMSTAR, there were still many deficiencies in MAs. We cannot attribute this improvement to the publication of PRISMA score and AMSTAR score, and there were other factors such as increased funding source. We expected that authors, journal editors and readers could pay more attention on the quality of MAs, to make the meta-analysis better serves the clinical work.

Role of the funding source

None.

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgments

We thank to the authors of the primary studies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.otsr.2017.08.009.

References