INTRODUCTION

Systematic reviews and meta-analyses aggregate the findings of multiple published works while taking into account the quality of evidence of each individual source.\(^1\) In addition, meta-analysis offers the benefit of combining data from individual studies into a larger sample, providing greater power to detect important differences in outcomes of interest.\(^2,3\) Because of this, systematic reviews and meta-analyses are generally regarded as important components of the collective body of orthopedic literature and are frequently utilized for clinical practice guidelines and recommendations.\(^4-6\)

However, performance of a quality systematic review or meta-analysis is contingent on many factors. Vital components of the process include appropriate study selection, quantitative and/or qualitative analysis, and data reporting. The methodological and reporting quality of these studies is key to their ability to generate unbiased results.\(^7\) Performed poorly, these of studies may lead to inaccurate conclusions.\(^8,9\) Because of this, it is important for authors to understand and adhere to available quality guidelines, such as the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and the Cochrane guidelines,\(^7,10,11\)

The issue of methodological quality in systematic reviews and meta-analyses has been explored previously in the orthopedic literature. In 2001, Bhandari et al. reviewed 40 meta-analyses published between 1969 and 1999 and found that the majority had major methodological flaws.\(^1\) Similarly, in 2010 Dijkman et al. reviewed orthopedic meta-analyses published during the years 2005 and 2008, and they concluded that a large percentage had major to extensive
More recently, in 2013, Gagnier et al. performed a review of systematic reviews and meta-analyses in the five highest impact factor orthopedic journals, utilizing the PRISMA checklist and the Assessment of Multiple Systematic Reviews (AMSTAR) index. They found that on average 68% of PRISMA checklist items and 54% of AMSTAR criteria were present, leading the authors to conclude, that the overall methodological and reporting quality remained “less than optimal.” These studies all highlights the necessity of strict adherence to appropriate methodological construct to ensure the quality of orthopedic systematic reviews and meta-analyses. Their findings are especially pertinent in light of the increasing number of systematic reviews and meta-analyses being published (an increase of 67% and 132%, respectively, between 2010 and 2014, compared to only a 27% increase for all PubMed-indexed items).

However, it is uncertain whether these prior studies, and similar works across other medical disciplines, have impacted the quality of systematic reviews and meta-analyses in the orthopedic literature. Furthermore, while providing recommendations to authors, they failed to offer guidance for readers and reviewers evaluating systematic reviews and meta-analyses. If the quality of studies being produced remains variable, it is incumbent on readers and reviewers to be able to critically evaluate quality when considering a study’s conclusions. The aim of our study is to examine methodological and reporting quality of orthopedic systematic reviews and meta-analyses published in 2016, across a range of orthopedic subspecialties, and recommend a quality rating strategy for readers and reviewers.

METHODS

We drafted a protocol outlining a search, data collection, and analysis strategy before initiation of our literature search. This protocol was registered on PROSPERO, an international registry of systematic review protocols. We conducted our review utilizing standard methodology as outlined in the Cochrane Handbook and reported results in accordance with the Preferred PRISMA guidelines.

Search strategy

One subspecialty society was chosen by faculty at our institution to represent each of the 10 orthopedic surgery subspecialties with accredited fellowships in the US, as outlined by the Accreditation Council for Graduate Medical Education (ACGME). Shoulder and elbow surgery, which is ACGME accredited under the classification of orthopedic sports medicine and adult reconstructive surgery, was included as a distinct subspecialty. The affiliated journals of each chosen subspecialty society were identified: Orthopedic Trauma (Journal of Orthopedic Trauma); Orthopedic Sports Medicine (The American Journal of Sports Medicine); Hand Surgery (Journal of Hand Surgery); Adult Reconstructive Orthopedic Surgery (The Journal of Arthroplasty); Orthopedic Surgery of the Spine (The Spine Journal); Shoulder and Elbow Surgery (Journal of Shoulder and Elbow Surgery); Pediatric Orthopedic Surgery (Journal of Pediatric Orthopedics); and Foot and Ankle Orthopedic Surgery (Foot and Ankle International); Musculoskeletal Oncology (Clinical Orthopedics and Related Research). If a subspeciality society had more than one affiliated journal, the journal with the highest impact factor, per the 2015 Thomson Reuters ISI index, was chosen. In addition, the highest impact orthopedic surgery journal (The Journal of Bone and Joint Surgery) was reviewed and included in our analysis.

Online archives of selected journals were searched for articles published during the calendar year 2016. Initial search was performed by reviewing the title and abstract for each study. To ensure completeness of study identification, a search of Medline was also performed using the keywords “systematic review OR meta-analysis” and “journal name OR journal abbreviation” and “2016.” All articles that stated in the title or methodology section that they were a systematic review and/or meta-analysis were identified. Full text of identified articles was subsequently reviewed to ensure that the article met inclusion criteria. Because our study aimed specifically to evaluate published literature, no attempts were made to obtain unpublished manuscripts.

To be included in the study, an article must have been explicitly described as a systematic review and/or a meta-analysis (including the terms “systematic review” and/or “meta-analysis”) in the title, introduction, or methods section. Excluded from analysis were published abstracts for which full-text manuscripts, published in 2016, were not available. Studies were also excluded if they did not include a formal search and/or data analysis strategy (i.e. expert opinion literature reviews). Eligibility criteria for each study were reviewed by two authors (STK and MJS). Disagreements between reviewers were resolved by consensus.

Data extraction and assessment of methodological quality

All data extraction was performed by two authors (STK and MJS) utilizing a uniform data extraction worksheet. Extracted data included primary author, author’s hospital and/or institutional affiliation, journal of publication, number of primary studies included, type of study designs included, reported level of evidence, protocol registration, financial support, and conflicts of interest reporting. Methodological and reporting quality was assessed by two authors (STK and MJS) using the guidelines in the PRISMA statement, as well as the AMSTAR scale. Disagreements between reviewers were resolved by consensus.

The PRISMA statement is an evidenced-based guideline for the performance and reporting of systematic reviews and meta-analyses.
It consists of a 27-item checklist (composed of seven main categories) and a flow chart. A PRISMA score (1–27) was assigned by reviewers based on presence or absence of each element within the paper. Item number 5 on the PRISMA checklist, relating to protocol, and registration was scored “yes” if the study explicitly described a protocol, regardless of whether or not the protocol was published. Information on protocol registration was captured as a separate data item on our uniform data collection worksheet.

The AMSTAR tool is a validated 11-item assessment of methodological quality that generates a score of (0–11). AMSTAR scores were assigned by reviewers based on presence or absence of each criteria. Disagreements between reviewers were resolved by consensus. To avoid ambiguity with regard to the interpretation of AMSTAR item number 5, we chose to grade this item as “yes” if studies excluded at the full-text review stage were referenced, consistent with the recommendation of prior authors who have critically evaluated the AMSTAR.

The AMSTAR checklist items are provided in Figure 1.

**Data analysis**

Because some items in both the PRISMA and AMSTAR pertain to quantitative analysis of data and may be relevant only to meta-analyses, we analyzed our data both in aggregate and separately for systematic reviews (studies containing only qualitative analysis) and meta-analyses (studies containing quantitative analysis). Distribution of PRISMA checklist items and AMSTAR scores was described using mean, standard deviation, and range. We also categorized studies based on the relative proportion of PRISMA and AMSTAR items present, to define a quality rating scale corresponding to the interquartile range for each metric. Interrater reliability was assessed using kappa. Correlation of PRISMA and AMSTAR scores for our aggregate sample was assessed using Pearson’s correlation. One-way analysis of variance (ANOVA) was used to evaluate the association of journal of publication with PRISMA and AMSTAR scores. Statistical significance was set at $P < 0.05$. Statistical analysis was performed using Stata Version 14 (StataCorp. College Station, TX: StatCorp LP, 2016).

**RESULTS**

Initial search results yielded 104 papers [Figure 2]. Full-text review excluded seven articles, leaving 97 articles for inclusion in the study (46 systematic reviews and 51 meta-analyses) [Appendix]. Description of studies by journal of publication is provided in Table 1. Only 43% (42/97) of studies reported level of evidence, with reported level of evidence ranging from 2 to 4 for systematic reviews and 1–4 for meta-analyses. Twelve percent (12/97) of studies were registered on PROSPERO (nine systematic reviews and three meta-analyses). Interrater agreement was strong for both PRISMA ($k = 0.85$ [95% confidence interval [CI] 0.82–0.88], $P < 0.001$) and AMSTAR ($k = 0.94$ [95% CI 0.90–0.98], $P < 0.001$). Correlation between PRISMA score and AMSTAR scores was also strong ($r = 0.83$, $P < 0.001$) [Figure 3]. Mean PRISMA score for our aggregate sample of systematic reviews and meta-analyses was 21 (±3.7, range 9–27), and mean AMSTAR score was 5.3 (±2.3, range 0–10). Itemized distribution of AMSTAR scoring for systematic reviews and meta-analyses is provided in Figure 4.

Among the 46 systematic reviews, 63% (29/46) provided detailed descriptions of included studies. Of these 29, 55% (16/29) included randomized control trials, 45% (13/29) included prospective cohort studies, 62% (18/29) included retrospective cohort studies, and 41% (12/29) included case series. Only 17% (5/29) included randomized control trials exclusively. For systematic reviews, the mean PRISMA score was 18.6 (±3.2, range 9–27). The lowest quartile of PRISMA
Table 1: Description of the study sample

<table>
<thead>
<tr>
<th>Subspecialty</th>
<th>Journal</th>
<th>Impact factor*</th>
<th>Number of systematic reviews meeting eligibility criteria</th>
<th>Number of meta-analyses meeting eligibility criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Orthopedics</td>
<td>The Journal of Bone and Joint Surgery</td>
<td>5.163</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Adult Reconstructive Orthopedic Surgery</td>
<td>The Journal of Arthroplasty</td>
<td>2.515</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Orthopedic Trauma</td>
<td>Journal of Orthopedic Trauma</td>
<td>1.850</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Shoulder and Elbow Surgery</td>
<td>Journal of Shoulder and Elbow Surgery</td>
<td>2.412</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Hand Surgery</td>
<td>Journal of Hand Surgery</td>
<td>1.640</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Orthopedic Surgery of the Spine</td>
<td>The Spine Journal</td>
<td>2.660</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Pediatric Orthopedic Surgery</td>
<td>Journal of Pediatric Orthopedics</td>
<td>1.330</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Musculoskeletal Oncology</td>
<td>Clinical Orthopedics and Related Research</td>
<td>3.127</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Foot and Ankle Orthopedic Surgery</td>
<td>Foot and Ankle International</td>
<td>1.896</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

*2015 Thomson Reuters ISI index

Figure 3: Correlation of preferred reporting items for systematic reviews and meta-analyses and assessment of multiple systematic reviews scores for all studies

Scores was 9–16, the interquartile range was 17–20, and the highest quartile was 21–27. The mean AMSTAR score was 4.2 (±2.2, range 0–9). The lowest quartile of AMSTAR scores was 0–1, the interquartile range was 2–6, and the highest quartile was 6–9. ANOVA demonstrated that mean scores

Figure 4: Distribution of assessment of multiple systematic reviews scores for systematic reviews and meta-analyses did not vary significantly by journal of publication for either PRISMA scores ($F(9,36) = 0.95$, $P = 0.497$) or AMSTAR scores ($F(9,36) = 1.12$, $P = 0.376$).
Table 2: Mean PRISMA and AMSTAR Scores by Journal of Publication

<table>
<thead>
<tr>
<th>Journal</th>
<th>Systematic Reviews</th>
<th>Meta-Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRISMA</td>
<td>AMSTAR</td>
</tr>
<tr>
<td>The Journal of Bone and Joint Surgery</td>
<td>20.5 (2.1)</td>
<td>5.5 (0.7)</td>
</tr>
<tr>
<td>The Journal of Arthroplasty</td>
<td>21.8 (4.9)</td>
<td>6.5 (2.9)</td>
</tr>
<tr>
<td>Journal of Orthopedic Trauma</td>
<td>17 (-)</td>
<td>4 (-)</td>
</tr>
<tr>
<td>The American Journal of Sports Medicine</td>
<td>18.6 (2.2)</td>
<td>3.8 (1.9)</td>
</tr>
<tr>
<td>Journal of Shoulder and Elbow Surgery</td>
<td>19.3 (3.0)</td>
<td>3.7 (2.1)</td>
</tr>
<tr>
<td>Journal of Hand Surgery</td>
<td>19 (-)</td>
<td>2 (-)</td>
</tr>
<tr>
<td>The Spine Journal</td>
<td>18.3 (3.0)</td>
<td>4.6 (2.6)</td>
</tr>
<tr>
<td>Journal of Pediatric Orthopedics</td>
<td>18.0 (3.2)</td>
<td>3.5 (1.7)</td>
</tr>
<tr>
<td>Clinical Orthopedics and Related Research</td>
<td>17 (1.7)</td>
<td>4.7 (1.5)</td>
</tr>
<tr>
<td>Foot and Ankle International</td>
<td>15 (8.5)</td>
<td>2 (0)</td>
</tr>
</tbody>
</table>

Mean (standard deviation). -: Indicates there were no studies included in that category. --: Indicates there was not standard deviation because there was only one study included in that category. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; AMSTAR: Assessment of Multiple Systematic Reviews.

Among the 51 meta-analyses, 82% (42/51) provided detailed descriptions of included studies. Of these 42, 80% (33/42) included randomized control trials, 56% (23/42) included prospective cohort studies, 54% (22/42) included retrospective cohort studies, and 20% (8/42) included case series. Only 25% (13/42) included randomized control trials exclusively. For meta-analyses, the mean PRISMA score was 23.3 (±2.7, range 16–27). The lowest quartile of PRISMA scores was 16–20, the inter-quartile range was 21–25, and the highest quartile was 26–27. The mean AMSTAR score was 6.4 (±1.9, range 2–10). The lowest quartile of AMSTAR scores was 2–4, the interquartile range was 5–7, and the highest quartile was 8–10. Mean PRISMA and AMSTAR scores by journal are provided in Table 2. ANOVA demonstrated that mean scores did not vary significantly by journal of publication for either PRISMA scores ($F (7,43) = 1.12, P = 0.366$) or AMSTAR scores ($F (7,43) = 2.12, P = 0.062$).

**DISCUSSION**

Our review demonstrates a wide range of methodological and reporting quality among included studies. Furthermore, these findings suggest that the overall quality of orthopedic systematic reviews and meta-analyses has not improved since this issue was first raised, nearly two decades ago.[15,12] Our findings are consistent with those reported by similar studies evaluating systematic reviews and meta-analyses in orthopedics and other areas of medicine.[13,20,36] Moreover, the scores reported in this study are similar to those published by Shea et al., for studies reviewed in both their initial publication and external validation studies of the AMSTAR, in which they reported ranges of scores from 0 to 10 and mean scores of 5.4 and 4.6, respectively.[18,19]

One important consideration when evaluating the findings of this study is interpretation of individual quality metric questions. Two specific items, of note, are AMSTAR questions 5 and 11. Gagnier et al., in their 2013, study reported 86% of studies fulfilled item 5, and 86% of studies fulfilled item 11; however, in our analysis, these items were fulfilled in far fewer studies (12% [12/97] for item 5 and 3% [3/97] for item 11).[13] With regard to AMSTAR item 5, we suggest interpretation of this criteria in accordance with the recommendations of Burda et al. and Wegewitz et al.[20,22] They suggest the item be interpreted such that it is graded as “yes” if studies excluded at the full-text review stage are referenced. We believe this item is important, as it relates to the transparency and reproducibility of the study selection process. AMSTAR item 11 asks specifically about inclusion of conflict of interest for the systematic review or meta-analysis authors and for the included studies. While most studies included a conflict of interest statement, very few referenced any potential conflicts of interest among the studies they included in their analysis. We believe this is also an important item, particularly if the systematic review or meta-analysis is including studies related to a product or service for which some of the existing published evidence may be biased by a pertinent conflict of interests.[18,22]

One of the more surprising findings of our analysis was that only 63% (29/46) of systematic reviews and 82% (42/51) of meta-analyses provided detailed descriptions of the types of studies included. This suggests a need for greater transparency regarding included studies, as it facilitates critical evaluation by readers and promotes a greater degree of confidence in the results of a systematic review or meta-analysis.[14] Furthermore, surprising was the large percentage of both systematic reviews (41%) and meta-analyses (20%)
which included case series in their analysis. This contrast with the relatively low percentage of systematic reviews (17%) and meta-analyses (25%) which included randomized control trials exclusively and suggests that the primary source data utilized in many of these studies may be suboptimal.\[37\]. In addition, our study found that 57% (55/97) of studies did not report level of evidence. Because both systematic reviews and meta-analyses can range in level of evidence (between 2–4 and 1–4 in our study, respectively), we recommend that all of these studies clearly state their level of evidence, as determined by the authors and/or editors of the publishing journal.

There are several limitations of the current study. This study does not provide an exhaustive assessment of all systematic reviews and meta-analyses published in all orthopedic journals. However, our process of study selection was designed to provide a broad overview, applicable to a wide range of orthopedic providers. As such, this work represents an important cross-sectional assessment of the current state of quality in this domain of literature. Another important consideration is that the quality of published systematic reviews and meta-analyses is partially dictated by the editorial process of the publishing journal and may vary among journals not included in our study. However, among our sample, journal of publication was not significantly associated with variation in either PRISMA or AMSTAR scores, despite the fact that the journals, we included covered a breadth of subspecialty interests and had a wide range of impact factors (1.330–5.163).\[17\] This supports the generalizability of our findings, and we believe that the 97 articles included in our study are likely representative of published orthopedic systematic reviews and meta-analyses.

**CONCLUSIONS**

Despite the stated limitations, our findings demonstrate that there remains substantial variability in methodological and reporting quality of systematic reviews and meta-analyses in the orthopedic literature. This suggests that prior calls for greater adherence to methodologic guidelines have been largely ineffective. The continued variability in quality, combined with the large volume of systematic reviews and meta-analyses being published, places a strong impetus on readers and reviewers to evaluate and formulate their own conclusions about the quality of these studies. To facilitate this process, we devised a strategy that allows assignment of a three-tier (low, moderate, or high) quality rating, using PRISMA and AMSTAR scores. Our findings indicate that a score ≥21 PRISMA items and ≥6 AMSTAR items for systematic reviews and ≥26 PRISMA items and ≥8 AMSTAR items for meta-analyses are associated with the top quartile of quality reporting and methodology. A score ≤16 PRISMA items and ≤1 AMSTAR items for systematic reviews and ≤20 PRISMA items and ≤4 AMSTAR items for meta-analyses is associated with the lowest quartile. Incorporating these score ranges into a single metric, we produced a simple grading scale [Table 3] by which readers and reviewers can critically assess and communicate study quality for orthopedic systematic reviews and meta-analyses.

Studies are considered a “systematic review” if they include only qualitative analysis of primary sources, studies are considered a “meta-analysis” if they included quantitative analysis primary sources. Quality grade is determined by the lowest scored metric.

**REFERENCES**

9. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 2016;94:485-514.
17. Reuters T. 2015 ISI Thomson Reuters Journal Citation Reports; 2016.