With the rapid proliferation of traditional printed and Internet-based literature, it is becoming increasingly difficult for neurosurgeons to stay up to date without having to spend hours searching and reading articles. Thus, review articles play an important role in continuing medical education. Of the 3 types of review articles, a narrative or journalistic review is a

Methodology and reporting of meta-analyses in the neurosurgical literature

A review

PAUL KLIRO JR., M.D., M.P.H., CLINTON J. THOMPSON, M.S., BRIAN T. RAGEL, M.D., and FREDERICK A. BOOP, M.D.

Semmes-Murphy Neurologic & Spine Institute; Department of Neurosurgery, University of Tennessee Health Science Center, Memphis, Tennessee; School of Public Health and Health Services, The George Washington University, Washington, DC; and Department of Neurosurgery, Oregon Health & Science University, Portland, Oregon

Object. Neurosurgeons are inundated with vast amounts of new clinical research on a daily basis, making it difficult and time-consuming to keep up with the latest literature. Meta-analysis is an extension of a systematic review that employs statistical techniques to pool the data from the literature in order to calculate a cumulative effect size. This is done to answer a clearly defined a priori question. Despite their increasing popularity in the neurosurgery literature, meta-analyses have not been scrutinized in terms of reporting and methodology.

Methods. The authors performed a literature search using PubMed/MEDLINE to locate all meta-analyses that have been published in the JNS Publishing Group journals (Journal of Neurosurgery, Journal of Neurosurgery: Pediatrics, Journal of Neurosurgery: Spine, and Neurosurgical Focus) or Neurosurgery. Accepted checklists for reporting (PRISMA) and methodology (AMSTAR) were applied to each meta-analysis, and the number of items within each checklist that were satisfactorily fulfilled was recorded. The authors sought to answer 4 specific questions: Are meta-analyses improving 1) with time; 2) when the study met their definition of a meta-analysis; 3) when clinicians collaborated with a potential expert in meta-analysis; and 4) when the meta-analysis was the only focus of the paper?

Results. Seventy-two meta-analyses were published in the JNS Publishing Group journals and Neurosurgery between 1990 and 2012. The number of published meta-analyses has increased dramatically in the last several years. The most common topics were vascular, and most were based on observational studies. Only 11 papers were prepared using an established checklist. The average AMSTAR and PRISMA scores (proportion of items satisfactorily fulfilled divided by the total number of eligible items in the respective instrument) were 31% and 55%, respectively. Major deficiencies were identified, including the lack of a comprehensive search strategy, study selection and data extraction, assessment of heterogeneity, publication bias, and study quality. Almost one-third of the papers did not meet our basic definition of a meta-analysis. The quality of reporting and methodology was better 1) when the study met our definition of a meta-analysis; 2) when one or more of the authors had experience or expertise in conducting a meta-analysis; 3) when the meta-analysis was not conducted alongside an evaluation of the authors’ own data; and 4) in more recent studies.

Conclusions. Reporting and methodology of meta-analyses in the neurosurgery literature is excessively variable and overall poor. As these papers are being published with increasing frequency, neurological journals need to adopt a clear definition of a meta-analysis and insist that they be created using checklists for both reporting and methodology. Standardization will ensure high-quality publications.

Key Words • meta-analysis • quality • methodology • neurosurgery • reporting • AMSTAR • PRISMA

Abbreviations used in this paper: AMSTAR = Assessment of Multiple Systematic Reviews; CONSORT = Consolidated Standards of Reporting Trials; JNS = Journal of Neurosurgery; MOOSE = Meta-Analyis of Observational Studies in Epidemiology; OQAQ = Overview of Quality Assessment Questionnaire; PRISMA = Preferred Reporting Items for Systematic reviews and Meta-analyses; QUORUM = Quality of Reporting of Meta-analyses.
simple, broad review of a given topic, typically done by an expert in the field, often with no specified intervention or outcome(s) evaluated. A systematic review is a much more rigorous and labor-intensive undertaking that seeks to answer a clearly defined question through a multistep process that is carried out in accordance with explicit, transparent, and reproducible methods. Systematic reviews summarize findings from papers located using a thorough search strategy and meeting certain eligibility criteria. The purpose of conducting a systematic review in such a stepwise, thorough fashion is to limit the introduction of bias—any process that systematically and nonrandomly causes a deviation of results and inferences from the truth—thus making the conclusions of the review more reliable and accurate.45

A meta-analysis is a direct extension of and often included in a systematic review. It uses the same steps but integrates the data from the studies using statistical techniques to obtain a pooled-effect estimate with greater power to detect true differences (and, by definition, a reduced chance of false-negative results, or Type II error) and thus is less influenced by the findings of any one study. Meta-analyses have become increasingly popular in the medical literature but no less controversial.33 For example, there are cases in which the findings of meta-analyses that were based on small trials were subsequently disproven by a single large randomized trial.35,62 As with systematic reviews, strict adherence to methodology is critical with meta-analyses to make them valid and precise and to avoid the oft-stated “garbage-in, garbage-out” criticism.35

A major advance in evidence-based medicine has been the development of structured checklists for methodology and reporting of systematic reviews and meta-analyses. The first checklist specific to meta-analyses was the Quality of Reporting of Meta-analyses (QUOROM),58 which was published in 1999 and designed to address the suboptimal reporting of meta-analyses. QUOROM is similar to the Consolidated Standards of Reporting Trials (CONSORT, www.consort-statement.org) for reporting of randomized controlled trials, which was published in 1996.1420 In 2009, QUOROM was revised to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA, www.prisma-statement.org) to encompass both systematic reviews and meta-analyses and address several conceptual and practical advances in the science of secondary research.69 Because of the increasing number of published meta-analyses using observational studies, the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group created a 6-section reporting cookbook: background, search strategy, methods, results, discussion, and conclusion.100 The Cochrane Collaboration provides a detailed, standardized handbook for systematic reviews that must be used by all submitting authors (http://www.cochrane-handbook.org/).

Adequate reporting does not necessarily ensure that the contents of the document are valid and precise. In other words, a systematic review or meta-analysis may be well written, but if the methods employed to derive a pooled estimate of the effect size are flawed, then the entire process is of questionable value. Indeed, the creators were clear that PRISMA could be useful for critical appraisal of published systematic reviews and meta-analyses but should not be used as an instrument to gauge their quality.69 The overall quality of a systematic review or meta-analysis, in our opinion, is a function of proper reporting but, much more importantly, of using the correct methodology to limit bias and ensure the internal validity of the findings. Similar to checklists for reporting, checklists for methodology have been created. Several well-known early questionnaires were the Overview of Quality Assessment Questionnaire (OQAQ),76,77 the Potsdam guidelines,22 and the Sacks instrument.88 In 2007, the Assessment of Multiple Systematic Reviews (AMSTAR) was developed by combining elements of the OQAQ and the Sacks instrument as well as other items based on methodological advances that had been made since OQAQ and Sacks had been introduced.92 It has been found to be valid, feasible, reliable, and to have good inter-rater agreement.91,93 Since their development, PRISMA and AMSTAR have become widely accepted by many journals as the tools to ensure proper reporting and methodology of systematic reviews and meta-analyses. In addition to these checklists, many textbooks have been published that thoroughly describe the various components and their function within a meta-analysis or systematic review.18,34,404,112

It is our impression that the term “meta-analysis” is being used with increasing frequency in the neurosurgical literature; however, is it being used appropriately? Is meta-analysis being performed too frequently? The word seems to be used to denote any attempt at combining data from two or more publications. Until very recently, academic neurosurgery has not mandated the use of the various checklists to ensure quality of reporting and methodology. The journal Neurosurgery now requires that all meta-analyses be reported as recommended by PRISMA or MOOSE guidelines, but does not require the use of a methodology questionnaire.11 We set out to apply the PRISMA and AMSTAR checklists to all meta-analyses in the neurosurgical literature to assess the quality of reporting and methodology.

**Methods**

**Search**

A search was conducted to identify all meta-analyses that had ever been published in the JNS (Journal of Neurosurgery) Publishing Group journals (Journal of Neurosurgery, Journal of Neurosurgery: Spine, Journal of Neurosurgery: Pediatrics, and Neurosurgical Focus) or in Neurosurgery through 2012. The words “meta-analysis,” “metaanalysis,” “meta-analyses,” and “meta-analyses” were searched either in the title, abstract, or key words. The search was conducted on PubMed/MEDLINE, and the results were cross-referenced with searches performed on the web sites of the respective journals.

**Definition**

There are 6 components that we deemed necessary for a meta-analysis. We defined a meta-analysis as a:

1. multistep process that
2. seeks to answer a clearly defined question, or questions, by comparing the impact of one or more interventions, exposures, or possible risk or predictive factors (there must be an exposure group[s] and control group[s]) on prespecified outcome(s) by

3. combining the results of previously published randomized controlled trials or observational studies that were found by a

4. reproducible and comprehensive search strategy that met prespecified entry criteria.

5. An overall or pooled effect size (for example, odds ratio, rate ratio) is calculated by

6. using appropriate statistical computational methods (fixed- vs random-effects models, measures of heterogeneity or inconsistency).

Although important and preferred, we did not define subgroup or sensitivity analysis or an assessment of publication bias as required elements of a meta-analysis.

Meta-analyses were categorized as analytical, exploratory, or both, as defined by Anello and Fleiss. An analytical type was one in which the main goal was to determine the common or pooled estimate of the effect size. When the focus of the meta-analysis was to explain why the effect sizes vary (that is, to search of sources of variability such as subgroup or sensitivity analysis, meta-regression analysis), then it was considered exploratory.

A “pure” meta-analysis was defined as one in which the entire focus of the paper is the meta-analysis. Conversely, a meta-analysis where the authors first detail their own institution’s experience and then secondarily add their results to the literature and perform a meta-analysis was considered “mixed.”

An observational study, as defined by Peipert and Phipps, is an etiological or effectiveness study using cross-sectional study, a case series, a case-control design, a design with historical controls, or a cohort design.

**Data Collection**

AMSTAR and PRISMA were applied to all meta-analyses found in our search. AMSTAR is an 11-item document used to assess quality of methodology for systematic reviews and meta-analyses (Appendix 1). Although the creators of AMSTAR had 4 answer options for each item—“yes,” “no,” “can’t answer,” and “not applicable”—we judged that a “can’t answer” response was equivalent to a “no.” To get a “yes” for an item, the meta-analysis must contain all the major components within that item, as detailed in Appendix 1.

PRISMA is a checklist for reporting that has 27 entries divided into 7 sections: title, abstract, introduction, methods, results, discussion, and funding (Appendix 2). The PRISMA authors have provided a detailed Explanation and Elaboration document that details the meaning and rationale for each checklist item. As with AMSTAR, we recorded one of 3 answers for each item: “yes,” “no,” “not applicable.” Some of the items in the checklist contain multiple components, so if at least half of them were met, then we gave that item a “yes.” If less than half of the components were met or a key component was lacking, the item was given a “no.” The PRISMA document also contains a 4-phase flow diagram that depicts the process of identifying, screening, determining eligibility, and presenting the reader with the final number of papers that were used in the systematic review or meta-analysis (Fig. 1). We recorded whether such a flow diagram was present in the meta-analysis or if the authors provided enough data to construct one.

Each meta-analysis was scrutinized using the AMSTAR and PRISMA checklists, and a score was assigned to the paper. The score was the proportion of items present (that is, the number of items that were given a “yes” response) divided by the number of items in the respective instrument. Each paper was reviewed on 3 separate occasions by the primary author (P.K.). Other data collected for each meta-analysis included the type of papers that were used in the meta-analysis (randomized controlled trials, observational studies, or both), the type of meta-analysis (pure vs mixed, analytical vs exploratory vs both), the topic of the meta-analysis (spine, trauma, vascular, infection prevention, functional/epilepsy, pain, oncology, peripheral nerve, pediatrics, and other), whether it satisfied our definition of a meta-analysis, whether there was one or more individuals identified as contributing to the paper who conceivably had additional training in meta-analysis techniques, such as a biostatistician, epidemiologist, or one with a master’s degree in public health, and whether the authors reported using one or more established reporting or methodology questionnaires as mentioned in the Introduction.

**Questions**

We sought to answer the following 4 questions: Is the quality of reporting and methodology of meta-analyses in the neurosurgery literature

1. Improving with time? Time was divided into the following 3 time periods: 1990–1999, 2000–2009, and 2010–2012. This corresponds with before and after the release of the QUOROM and PRISMA statements.

2. Better if the authors or other acknowledged contributors include one or more individuals who might have additional expertise in conducting a meta-analysis?

3. Different when comparing pure versus mixed meta-analysis?

4. Different in those meta-analyses that met our definition compared with those that did not?

**Statistical Analysis**

For each paper, an AMSTAR and PRISMA “score” was calculated. The responses for each item from both checklists were collapsed into “yes” versus “not yes” categories, with the “yes” responses being summed then divided by the total number of items in the respective instrument. The number of items in the AMSTAR instrument—the denominator for the calculation of the score—is 11. The denominator used in the PRISMA score differed depending on whether the authors conducted any additional analyses (items 16 and 23 from the checklist). If they did not conduct any additional analyses (“not applicable”), then the denominator was reduced from 27 to 25 as a way to distinguish (and not penalize) those papers that did not include additional analyses from those that
The statistical analysis in the present paper was largely descriptive. Frequencies were calculated presenting the distribution of papers by journal, time period (1990–1999, 2000–2009, 2010–2012), pure versus mixed analysis, presence of meta-analysis expert, paper types constituting the meta-analyses, topic, and our assessment of whether the paper qualifies as a meta-analysis (yes vs no). Tests of statistical significance relied on the Wilcoxon rank-sum test or Spearman test, and significance was reached if \( p < 0.05 \). All \( p \) values are 2-sided.

Stata/SE 11.2 was used for all analyses.

**Results**

We identified 76 meta-analyses in the neurosurgery literature from 1990 through the end of 2012 using our search criteria (Table 1). Two articles were excluded from analysis because the meta-analysis performed was of such poor quality that AMSTAR and PRISMA could not be applied.\(^5,9\) The recent meta-analysis by van Leeuwen et al.\(^110\) was not analyzed because it is an individual patient data meta-analysis, and many of the items in both checklists were “N/A.” We also excluded our recent meta-analysis to avoid a conflict of interest.\(^5,8\) Thus, 72 papers were included in our analysis. The *Journal of Neurosurgery* and the other JNS Publishing Group journals published 65% of the meta-analyses, and the most common topics in all journals were vascular (21 meta-analyses, 29%). More than half of the meta-analyses (39, 54%) were based solely on observational studies and were either analytical (29, 40%) or analytical and exploratory (21, 29%). In 11 (15%) studies, the authors stated that they used one of the 3 available reporting checklists (MOOSE, QUOROM, PRISMA) in preparing their paper, but no one reported using a methodology checklist.

Figure 2 depicts the number of meta-analyses that were given a “yes” response to each of the 11 questions in AMSTAR. Fifty percent or more of the papers received a positive response for 3 items (Q1, Q6, Q9). Most papers clearly stated their research question and inclusion criteria (Q1), but there were very low responses for duplication of study selection and data extraction (Q2), use of “gray literature” (Q4), list of included and excluded studies (Q5), and assessment and analysis of study quality (Q7, Q8), and only one study documented source(s) of financial support for the meta-analysis and each of the included studies (Q11). Half of the papers used appropriate methods to combine findings (Q9). Publication bias was adequately assessed in 21 papers (29%, Q10).

The results for PRISMA questions are shown in Fig. 3. More than 50% of the papers (more than 36) had positive responses for 16 (59%) of the 27 PRISMA items. Conversely, there were low responses for the remaining 11 questions—29 or fewer papers successfully satisfied these items. High responses were recorded for the items in PRISMA that required using “meta-analysis” in the title of the paper, having a structured abstract, describing rationale...
TABLE 1: Characteristics of 72 meta-analyses evaluated

<table>
<thead>
<tr>
<th>Description</th>
<th>No. of Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>journal</td>
<td></td>
</tr>
<tr>
<td>JNS Publishing Group journals</td>
<td>47</td>
</tr>
<tr>
<td>Journal of Neurosurgery</td>
<td>35</td>
</tr>
<tr>
<td>Journal of Neurosurgery: Spine</td>
<td>8</td>
</tr>
<tr>
<td>Journal of Neurosurgery: Pediatrics</td>
<td>2</td>
</tr>
<tr>
<td>Neurosurgical Focus</td>
<td>2</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>25</td>
</tr>
<tr>
<td>topic</td>
<td></td>
</tr>
<tr>
<td>vascular</td>
<td>21</td>
</tr>
<tr>
<td>spine</td>
<td>13</td>
</tr>
<tr>
<td>oncology</td>
<td>9</td>
</tr>
<tr>
<td>trauma</td>
<td>8</td>
</tr>
<tr>
<td>infection prevention</td>
<td>7</td>
</tr>
<tr>
<td>pain</td>
<td>1</td>
</tr>
<tr>
<td>functional/epilepsy</td>
<td>6</td>
</tr>
<tr>
<td>pediatrics</td>
<td>2</td>
</tr>
<tr>
<td>peripheral nerve</td>
<td>1</td>
</tr>
<tr>
<td>other</td>
<td>4</td>
</tr>
<tr>
<td>types of studies</td>
<td></td>
</tr>
<tr>
<td>observational</td>
<td>39</td>
</tr>
<tr>
<td>randomized controlled trial</td>
<td>13</td>
</tr>
<tr>
<td>both</td>
<td>12</td>
</tr>
<tr>
<td>unknown/Neither/case series/case reports</td>
<td>8</td>
</tr>
<tr>
<td>type of meta-analysis</td>
<td></td>
</tr>
<tr>
<td>pure/mixed</td>
<td>58/14</td>
</tr>
<tr>
<td>analytic/exploratory/both/IPD*/NA†</td>
<td>29/12/12/19</td>
</tr>
</tbody>
</table>

* Analytic meta-analysis with individual patient data.
† Could not assign analytic or exploratory designation to the meta-analysis that did not meet our definition.

and objectives of the meta-analysis in the introduction, and providing inclusion/exclusion criteria and details of the search strategy, but few adequately reported the process for selecting studies and collecting data (Q9, Q10). No papers indicated that a protocol existed and/or was registered and accessible. As for AMSTAR, it was infrequent for a paper to assess bias within individual studies, that is, assigning some measure of quality to the included studies (Q12, Q19) and across studies, that is, publication bias (Q15, Q22). Almost all papers that adequately described methods for additional analyses, such as sensitivity analysis, subgroup analysis, or meta-regression, presented their results (Q16, Q23). The 3 items—summary of evidence, study limitations, and conclusions—for the Discussion (Q24–26) had positive responses in 44 or more of the papers. Funding, which unlike in the case of AMSTAR was specific only for the meta-analysis, was reported in 26 papers. Seventeen papers (24%) contained a search flow diagram or presented enough data to construct one.

Study Questions

The results of our 4 study questions are shown in
value of meta-analyses depends on the secondary research being conducted with sound methodology and being reported completely and with transparency. Reports that are biased or have flawed methodology can lead to inappropriate estimates of treatment effects.

We analyzed all qualifying meta-analyses published in the specified journals during the study period according to our search criteria, using AMSTAR and PRISMA for methodology and reporting, respectively. Overall, the mean "scores" for both checklists were less than ideal: 31% for AMSTAR and 53% for PRISMA. In other words, the average meta-analysis in the neurosurgery literature had satisfactorily fulfilled approximately 3 of 11 items in AMSTAR and 14 of 27 items in PRISMA. Thus, the average meta-analysis had better reporting than methodology. All but 2 of the domains in AMSTAR were fulfilled in less than half of the papers (Fig. 1). Only Items 1 (research question) and 6 (characteristics of included studies) were adequately completed. We found major deficiencies in the method of study selection and data extraction, which were identified through the PRISMA checklist.

### TABLE 2: AMSTAR and PRISMA results

<table>
<thead>
<tr>
<th>Description</th>
<th>AMSTAR</th>
<th>PRISMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall mean score (range)</td>
<td>31% (0–82%)</td>
<td>53% (4–93%)</td>
</tr>
<tr>
<td>time period (no. of papers, mean score)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990–1999</td>
<td>9, 19%</td>
<td>9, 35%</td>
</tr>
<tr>
<td>2000–2009</td>
<td>30, 31%</td>
<td>30, 52%</td>
</tr>
<tr>
<td>2010–2012</td>
<td>33, 34%</td>
<td>33, 63%</td>
</tr>
<tr>
<td>p = 0.1093</td>
<td>p = 0.0004</td>
<td></td>
</tr>
<tr>
<td>meta-analysis expert (no. of papers, mean score)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>31, 40%</td>
<td>31, 63%</td>
</tr>
<tr>
<td>no</td>
<td>41, 24%</td>
<td>41, 49%</td>
</tr>
<tr>
<td>p = 0.0003</td>
<td>p = 0.0117</td>
<td></td>
</tr>
<tr>
<td>pure vs mixed (no. of papers, mean score)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td>58, 35%</td>
<td>58, 59%</td>
</tr>
<tr>
<td>mixed</td>
<td>14, 16%</td>
<td>14, 38%</td>
</tr>
<tr>
<td>p = 0.0034</td>
<td>p = 0.0018</td>
<td></td>
</tr>
<tr>
<td>met our definition (no. of papers, mean score)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>51, 38%</td>
<td>51, 63%</td>
</tr>
<tr>
<td>no</td>
<td>21, 13%</td>
<td>21, 35%</td>
</tr>
<tr>
<td>p &lt; 0.0001</td>
<td>p &lt; 0.0001</td>
<td></td>
</tr>
</tbody>
</table>
of intracranial dural arteriovenous malformations is an example of what we contend is an inaccurate use of the term “meta-analysis.” The authors reviewed existing literature but simply summed the percentage of cases that were successes and failures for a given treatment modality and applied a chi-square test, which is not a proper method.36 The following lengthy commentary by one of the paper’s reviewers, an epidemiologist, summarizes precisely the problem with many meta-analyses in the neurosurgery literature (this commentary, along with others, was published along with the study):

We are concerned about describing this report as a “meta-analysis.” It is true that meta-analysis can be simply defined as a systematic review that uses statistical methods to combine and summarize the results of several trials, and any review of such as this might be argued to be a “meta-analysis.” However, epidemiologists have something different in mind when they discuss meta-analyses, or at least meta-analyses that are considered valid in the growing field of evidence-based medicine. The best meta-analyses compile information from controlled and/or randomized trials, but even when uncontrolled case series or observational studies (such as those presented in this review) are combined for statistical reanalysis, the reader has to have some reassurance that the patient populations receiving the different treatments are similar and homogeneous. In this review, for example, we really have no idea of what kind of patients are being compared, other than that their malformations were located in a certain anatomic region. No effort was made to present or adjust for patient characteristics that may have confounded the analysis, and it is not clear that the data collection was systematic or scored for quality. We are not sure that the conclusions of this combined analysis need to be considered any more valid than those of any one of the series included in its review.

Such “pseudo-meta-analyses” by Lucas et al. and others should be more precisely considered an attempt at conducting a systematic or narrative review. In our opinion, neurosurgeons, for the most part, have only a general understanding of meta-analyses and thus believe that they can be done with relative ease. This belief may be fostered by the easy availability and access to meta-analysis software such as RevMan (Cochrane Collaboration). To the contrary, however, Berman and Parker correctly state that a meta-analysis “…cannot be thought of as a quick and easy way to pull a lot of studies together and come up with a publication, but like any other study, requires an appreciable investment of time in planning and implementation.”

Although there is clearly overlap in the items between PRISMA and AMSTAR, PRISMA is only a guide on what to write in a meta-analysis paper; it does not provide the specifics on how to conduct the meta-analysis. More succinctly, adequacy of reporting does not ensure proper methodology. This is why the creators of PRISMA were clear in stating that it is “…not a quality assessment instrument to gauge the quality of a systematic review.” This point is exemplified by the meta-analysis by de Oliveira et al. The authors sought to answer whether clipping or coiling ruptured aneurysms had any effect on shunt dependency. This meta-analysis was fairly well reported, satisfying 14 of 25 PRISMA items (Q16 and Q23 were not applicable). The authors concluded that coiling was associated with a higher risk of shunt dependency than clipping, but they observed significant heterogeneity.

The meta-analysis by Lucas et al. on the treatment of intracranial dural arteriovenous malformations is an example of what we contend is an inaccurate use of the term “meta-analysis.” The authors reviewed existing literature but simply summed the percentage of cases that were successes and failures for a given treatment modality and applied a chi-square test, which is not a proper method. The following lengthy commentary by one of the paper’s reviewers, an epidemiologist, summarizes precisely the problem with many meta-analyses in the neurosurgery literature (this commentary, along with others, was published along with the study):

We are concerned about describing this report as a “meta-analysis.” It is true that meta-analysis can be simply defined as a systematic review that uses statistical methods to combine and summarize the results of several trials, and any review of such as this might be argued to be a “meta-analysis.” However, epidemiologists have something different in mind when they discuss meta-analyses, or at least meta-analyses that are considered valid in the growing field of evidence-based medicine. The best meta-analyses compile information from controlled and/or randomized trials, but even when uncontrolled case series or observational studies (such as those presented in this review) are combined for statistical reanalysis, the reader has to have some reassurance that the patient populations receiving the different treatments are similar and homogeneous. In this review, for example, we really have no idea of what kind of patients are being compared, other than that their malformations were located in a certain anatomic region. No effort was made to present or adjust for patient characteristics that may have confounded the analysis, and it is not clear that the data collection was systematic or scored for quality. We are not sure that the conclusions of this combined analysis need to be considered any more valid than those of any one of the series included in its review.

Such “pseudo-meta-analyses” by Lucas et al. and others should be more precisely considered an attempt at conducting a systematic or narrative review. In our opinion, neurosurgeons, for the most part, have only a general understanding of meta-analyses and thus believe that they can be done with relative ease. This belief may be fostered by the easy availability and access to meta-analysis software such as RevMan (Cochrane Collaboration). To the contrary, however, Berman and Parker correctly state that a meta-analysis “…cannot be thought of as a quick and easy way to pull a lot of studies together and come up with a publication, but like any other study, requires an appreciable investment of time in planning and implementation.”

Although there is clearly overlap in the items between PRISMA and AMSTAR, PRISMA is only a guide on what to write in a meta-analysis paper; it does not provide the specifics on how to conduct the meta-analysis. More succinctly, adequacy of reporting does not ensure proper methodology. This is why the creators of PRISMA were clear in stating that it is “…not a quality assessment instrument to gauge the quality of a systematic review.” This point is exemplified by the meta-analysis by de Oliveira et al. The authors sought to answer whether clipping or coiling ruptured aneurysms had any effect on shunt dependency. This meta-analysis was fairly well reported, satisfying 14 of 25 PRISMA items (Q16 and Q23 were not applicable). The authors concluded that coiling was associated with a higher risk of shunt dependency than clipping, but they observed significant heterogeneity.

The meta-analysis by Lucas et al. on the treatment of intracranial dural arteriovenous malformations is an example of what we contend is an inaccurate use of the term “meta-analysis.” The authors reviewed existing literature but simply summed the percentage of cases that were successes and failures for a given treatment modality and applied a chi-square test, which is not a proper method. The following lengthy commentary by one of the paper’s reviewers, an epidemiologist, summarizes precisely the problem with many meta-analyses in the neurosurgery literature (this commentary, along with others, was published along with the study):

We are concerned about describing this report as a “meta-analysis.” It is true that meta-analysis can be simply defined as a systematic review that uses statistical methods to combine and summarize the results of several trials, and any review of such as this might be argued to be a “meta-analysis.” However, epidemiologists have something different in mind when they discuss meta-analyses, or at least meta-analyses that are considered valid in the growing field of evidence-based medicine. The best meta-analyses compile information from controlled and/or randomized trials, but even when uncontrolled case series or observational studies (such as those presented in this review) are combined for statistical reanalysis, the reader has to have some reassurance that the patient populations receiving the different treatments are similar and homogeneous. In this review, for example, we really have no idea of what kind of patients are being compared, other than that their malformations were located in a certain anatomic region. No effort was made to present or adjust for patient characteristics that may have confounded the analysis, and it is not clear that the data collection was systematic or scored for quality. We are not sure that the conclusions of this combined analysis need to be considered any more valid than those of any one of the series included in its review.

Such “pseudo-meta-analyses” by Lucas et al. and others should be more precisely considered an attempt at conducting a systematic or narrative review. In our opinion, neurosurgeons, for the most part, have only a general understanding of meta-analyses and thus believe that they can be done with relative ease. This belief may be fostered by the easy availability and access to meta-analysis software such as RevMan (Cochrane Collaboration). To the contrary, however, Berman and Parker correctly state that a meta-analysis “…cannot be thought of as a quick and easy way to pull a lot of studies together and come up with a publication, but like any other study, requires an appreciable investment of time in planning and implementation.”

Although there is clearly overlap in the items between PRISMA and AMSTAR, PRISMA is only a guide on what to write in a meta-analysis paper; it does not provide the specifics on how to conduct the meta-analysis. More succinctly, adequacy of reporting does not ensure proper methodology. This is why the creators of PRISMA were clear in stating that it is “…not a quality assessment instrument to gauge the quality of a systematic review.” This point is exemplified by the meta-analysis by de Oliveira et al. The authors sought to answer whether clipping or coiling ruptured aneurysms had any effect on shunt dependency. This meta-analysis was fairly well reported, satisfying 14 of 25 PRISMA items (Q16 and Q23 were not applicable). The authors concluded that coiling was associated with a higher risk of shunt dependency than clipping, but they observed significant heterogeneity.

The meta-analysis by Lucas et al. on the treatment of intracranial dural arteriovenous malformations is an example of what we contend is an inaccurate use of the term “meta-analysis.” The authors reviewed existing literature but simply summed the percentage of cases that were successes and failures for a given treatment modality and applied a chi-square test, which is not a proper method. The following lengthy commentary by one of the paper’s reviewers, an epidemiologist, summarizes precisely the problem with many meta-analyses in the neurosurgery literature (this commentary, along with others, was published along with the study):

We are concerned about describing this report as a “meta-analysis.” It is true that meta-analysis can be simply defined as a systematic review that uses statistical methods to combine and summarize the results of several trials, and any review of such as this might be argued to be a “meta-analysis.” However, epidemiologists have something different in mind when they discuss meta-analyses, or at least meta-analyses that are considered valid in the growing field of evidence-based medicine. The best meta-analyses compile information from controlled and/or randomized trials, but even when uncontrolled case series or observational studies (such as those presented in this review) are combined for statistical reanalysis, the reader has to have some reassurance that the patient populations receiving the different treatments are similar and homogeneous. In this review, for example, we really have no idea of what kind of patients are being compared, other than that their malformations were located in a certain anatomic region. No effort was made to present or adjust for patient characteristics that may have confounded the analysis, and it is not clear that the data collection was systematic or scored for quality. We are not sure that the conclusions of this combined analysis need to be considered any more valid than those of any one of the series included in its review.

Such “pseudo-meta-analyses” by Lucas et al. and others should be more precisely considered an attempt at conducting a systematic or narrative review. In our opinion, neurosurgeons, for the most part, have only a general understanding of meta-analyses and thus believe that they can be done with relative ease. This belief may be fostered by the easy availability and access to meta-analysis software such as RevMan (Cochrane Collaboration). To the contrary, however, Berman and Parker correctly state that a meta-analysis “…cannot be thought of as a quick and easy way to pull a lot of studies together and come up with a publication, but like any other study, requires an appreciable investment of time in planning and implementation.”

Although there is clearly overlap in the items between PRISMA and AMSTAR, PRISMA is only a guide on what to write in a meta-analysis paper; it does not provide the specifics on how to conduct the meta-analysis. More succinctly, adequacy of reporting does not ensure proper methodology. This is why the creators of PRISMA were clear in stating that it is “…not a quality assessment instrument to gauge the quality of a systematic review.” This point is exemplified by the meta-analysis by de Oliveira et al. The authors sought to answer whether clipping or coiling ruptured aneurysms had any effect on shunt dependency. This meta-analysis was fairly well reported, satisfying 14 of 25 PRISMA items (Q16 and Q23 were not applicable). The authors concluded that coiling was associated with a higher risk of shunt dependency than clipping, but they observed significant heterogeneity.
(as indicated from the test of heterogeneity), which should have prompted them to reconsider use of a fixed-effects model. A random-effects model, in addition to being a more appropriate method, also had the drastic effect of nullifying their conclusion: there was no longer a benefit to surgery compared with coiling.\textsuperscript{81}

Despite our strong criticisms, we have demonstrated that meta-analyses in the neurosurgery literature do seem to be improving in both reporting and methodology with time. They are also of higher quality when the authors focused solely on designing and conducting a meta-analysis and when the authors included one or more persons that had specialized training in conducting a meta-analysis. It has been argued that the creation of a meta-analysis requires a team, which should include both statisticians and physicians, each bringing their own area of expertise to the task at hand.\textsuperscript{45,112}

**Literature Review**

Although our study represents the first attempt, to our knowledge, to apply the PRISMA checklists to existing meta-analyses, retrospective assessment of methodology quality has been performed before. Delaney et al.\textsuperscript{27} used OQAQ to assess the quality of reports in the critical care literature before and after the publication of the QUOROM statement. They found that overall quality was poor (mean overall OQAQ score 3.3 of 7) and noted particular shortcomings in the literature search and avoidance of bias in the inclusion of studies, but they did find that the quality improved after the release of the QUOROM statement. Dixon et al.\textsuperscript{31} found that most studies in the general surgery literature, even after QUOROM, had major methodological flaws, with a median OQAQ score of 3.3. Factors associated with low overall scientific quality included the absence of any prior meta-analysis publications by the authors and meta-analyses produced by surgical department members without external collaboration, a finding that we observed as well. Dijkman et al.\textsuperscript{30} found that the quality of methodology of meta-analyses in the orthopedic literature did improve with time but that a substantial proportion continued to show extensive flaws, again similar to our results. Similar appraisals of quality of systematic reviews have been done in the anesthesiology,\textsuperscript{29} general medical,\textsuperscript{66} and emergency medicine\textsuperscript{55} literature, as well as a cross-sectional sample of all systematic reviews entered into MEDLINE at a particular time and year.\textsuperscript{71} Turner et al.\textsuperscript{107} recently demonstrated that trials published in journals that endorse the CONSORT checklist are significantly more completely reported for several key CONSORT items than those published in nonendorsing journals or prior to endorsement in a given journal. Kiehn et al. found that the mean CONSORT scores were higher in 3 major medical journals (JAMA, Lancet, New England Journal of Medicine) that endorsed the CONSORT statement than in 5 neurosurgical journals, including Journal of Neurosurgery and Neurosurgery, that did not (41 vs 26.4).\textsuperscript{56}

**Is a Meta-Analysis Appropriate?**

In the process of formulating the primary question for a meta-analysis, the researcher must also ask whether the question is clinically relevant and whether it is appropriate to compare groups of patients. For example, a recent well-conducted meta-analysis was designed to answer whether there was a difference in the improvement in leg pain and complication rates in patients undergoing a traditional open lumbar microdiscectomy compared with those who had a minimally invasive discectomy.\textsuperscript{24} There is little doubt that this constitutes a clinically relevant and valid comparison and is worthy of a meta-analysis; however, other meta-analyses have posed questions that are of questionable clinical relevance or inappropriate for a meta-analysis. For example, one could argue that most children who undergo a craniotomy for a craniofaryngioma have a larger and more complex tumor than those children who undergo a transsphenoidal operation, usually having a tumor confined within or just above the sella.\textsuperscript{37} Contrasting these 2 groups of patients is certainly clinically pertinent, but it is debatable whether a formal meta-analysis can or should be done. Other examples included evaluating the survival of patients with glioblastoma who underwent a biopsy only versus a resection,\textsuperscript{106} the timing of surgery for patient with aneurysmal subarachnoid hemorrhage,\textsuperscript{23} and comparing the risk of cranial neuropathy in patients who underwent radiosurgery compared with surgery for cavernous sinus meningiomas.\textsuperscript{101} In other papers, there was no comparison group at all.\textsuperscript{13,17,105,115} For these reports, the term “meta-analysis,” according to our definition, would not have been appropriate. Neurosurgery has recently stated that they will not accept noncomparative meta-analyses for publication because of the many undetectable and insurmountable biases.\textsuperscript{11} In summary, a meta-analysis should not be done if the question at hand is currently clinically irrelevant or outdated, if the treatment and control groups are so different as to preclude a meaningful comparison, or if there is no comparison group. In these situations, a systematic review of the literature may be the better choice.

**Study Limitations**

There are a number of limitations to the present study. Although we included a large number of meta-analyses in our evaluation, our search may not have identified all meta-analyses published in the specified neurosurgical journals. Some meta-analyses may also have been missed if they were contained within a systematic review and the authors did not identify it as a meta-analysis. Some of the items in the questionnaires, particularly PRISMA, introduce an element of subjectivity when there were multiple components within a single item (see Appendices 1 and 2). As stated in Methods, we were much more stringent in categorizing whether a paper fulfilled an item within AMSTAR than within PRISMA. Likewise, assigning an all-or-none score for an item may be misleading, falsely lowering the perceived quality of a paper. This has recently been addressed by Kung et al.,\textsuperscript{64} who created a revised AMSTAR (R-AMSTAR) in which each of the 11 items is subdivided into 4 components with a score ranging from 11 (where none of the items were fulfilled) to 44. This revised AMSTAR permits a detailed analysis of acceptable versus deficient aspects of each systematic review or meta-analysis and reliably quantifies the 11 domains;
however, R-AMSTAR has yet to be widely tested or accepted. Our use of the AMSTAR and PRISMA checklists to calculate a “score” is also a valid criticism and limitation since neither instrument was intended for use in this way. Our score, however, is a simple proportion defined as the number of items fulfilled in the instrument divided by the total number of items contained in the instrument. The respective scores are straightforward and easily interpretable ways to present and quantify the quality of a meta-analysis that has, also, precedent. Finally, there were a few meta-analyses of individual patient data. These represent rare and unique meta-analyses in which the application of PRISMA and AMSTAR is challenging and should be conducted using established techniques.

Our Recommendations

The results of our study underscore the need to standardize the way meta-analyses (and systematic reviews) are designed and reported. Applying PRISMA and AMSTAR to future meta-analyses should result in higher quality in both reporting and methodology and decrease the inappropriate and incorrect use of meta-analytical techniques (what we term “pseudo-meta-analyses”) and the word “meta-analysis.” We offer the following recommendations.

1. A clear definition of a meta-analysis (and systematic review) needs to be adopted.

2. If prospective authors decide to use the term “meta-analysis” to describe their study, both PRISMA and AMSTAR need to be applied. These checklists should be easily accessible through the website of the journal to which they wish to submit. If their study does not conform to the definition of a “meta-analysis” or “systematic review,” then other terms such as “narrative review” or “literature review” should be used instead, with the recognition that these studies will not meet the same criteria.

3. Reviewers for the journal need to be familiar or “trained” in these checklists so that they can recognize whether the items have been fulfilled.

4. Physicians should team up with one or more persons who have expertise in the statistical techniques of meta-analyses.

5. We encourage journals to create a link on their website where all systematic reviews and meta-analyses that have been published by that journal are categorized and readily accessible to readers. Prospective authors can view what meta-analyses have already been done, and neurosurgeons can rapidly find topic reviews to assist them in the application of evidence-based medicine to their clinical practice.

Conclusions

Oxman et al.77 stated, “The fact that a review article is published in a peer-reviewed journal, even a prestigious one, is no guarantee of scientific quality.” Our study verifies this claim. We have shown that there is wide variability in the reporting and methodology of meta-analyses in the neurosurgery literature by applying the PRISMA and AMSTAR checklists. Many meta-analyses contained major deficiencies and some used the word “meta-analysis” inappropriately or incorrectly. The overall quality, however, is improving with time and is better when the authors collaborate with others who are familiar with meta-analysis techniques. The publishers of the Journal of Neurosurgery, JNS Publishing Group, should insist that reporting and methodology checklists be used for every submitted meta-analysis to elevate the quality of these important pieces of evidence-based documents to levels on par with other medical journals.

Acknowledgment

We thank Kristin Kraus, M.Sc., for editorial assistance with the paper.

Disclosure

The authors do not report any conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author contributions to the study and manuscript preparation include the following. Concept and design: all authors. Acquisition of data: Klimo. Analysis and interpretation of data: Klimo, Thompson, Ragel. Drafting the article: all authors. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Klimo.

References

Methodology and reporting of neurosurgical meta-analyses
disk prostheses in clinical practice: a meta-analysis of ran-
domized controlled trials. 

13. Bauer DF, Razdan SN, Bartolucci AA, Markert JM: Meta-
analysis of hemorrhagic complications from ventriculostomy placement by neurosurgeons. 


19. Chappell ET, Mouré FC, Good MC: Comparison of computed tomographic angiography with digital subtraction angiogra-
phy in the diagnosis of cerebral aneurysms: a meta-analysis. 


25. de Gans K, Nieuwkamp DJ, Rinkel GJ, Algra A: Timing of 

sion of ruptured intracranial aneurysms by surgical clipping or endovascular coiling: a single-institution series and meta-

27. Delaney A, Baqashaw SM, Ferland A, Manns B, Laupland KB, Doig CJ: A systematic evaluation of the quality of meta-
analyses in the critical care literature. Crit Care 9:R575–R582, 2005


29. Di Maio S, Temkin N, Ramanathan D, Sekhar LN: Current 


35. Egger M, Smith GD, Sterne JA: Uses and abuses of meta-


37. Elliott RE, Jane JA Jr, Wisoff JH: Surgical management of cranioopharyngiomas in children: meta-analysis and com-


39. Englot DJ, Chang EF, Auguste KI: Vagus nerve stimulation for epilepsy: a meta-analysis of efficacy and predictors of re-

40. Englot DJ, Han SJ, Lawton MT, Chang EF: Predictors of seizure freedom in the surgical treatment of supratento-


42. Farace E, Alves WM: Do women fare worse: a metaanaly-

43. Fraser GF, Härtl R: Anterior approaches to fusion of the cervi-


45. Greenhalgh T: Papers that summarise other papers (system-

rior thoracic spinal cord herniation: three new cases and an individual patient data meta-analysis of 126 case reports.

47. Gross BA, Du R: The natural history of cerebral dural arterio-


49. Hamilton MG, Yee WH, Hull RD, Ghali WA: Venous thromboembolism prophylaxis in patients undergoing cranial neuro-


related morbidity for patients with glomus jugulare tumors. 

53. Jiraratnapochchai K, Jung S: Nonsteroidal antiinflammatory 
drugs for postoperative pain management after lumbar spine 
surgery: a meta-analysis of randomized controlled trials. J 

H: Anterior interosseous nerve compression after supracon- 
dylar fracture of the humerus: a metaanalysis. J Neurosurg 
90:1053–1056, 1999

55. Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH: Evaluating 
the utility of systematic reviews in the emergency 

56. Kiehna EN, Starke RM, Pouratian N, Dumont AS: Standards 
for reporting randomized controlled trials in neurosurgery. A 

from elective surgery for asymptomatic, unruptured, intracra- 
1994

58. Klimo P Jr, Thompson CJ, Ragel BT, Boop FA: Antibiotic- 
impregnated shunt systems versus standard shunt systems: a 
meta- and cost-savings analysis. Clinical article. J Neurosurg 
Pedia 8:600–612, 2011

59. Komotar RJ, Starke RM, Otten ML, Merkow MB, Garrett 
MC, Marshall RS, et al: The role of indirect extracranial-
intracranial bypass in the treatment of symptomatic intracra- 
nal atheroocclusive disease. Clinical article. J Neurosurg 
110:896–904, 2009

60. Krishna V, Kim DH: Ethnic differences in risk factors for sub- 

L, et al: From systematic reviews to clinical recommendations 
for evidence-based health care: validation of revised assess- 
ment of multiple systematic reviews (R-AMSTAR) for grad- 

F: Discrepancies between meta-analyses and subsequent large 
1997

63. Levy EI, Firlik AD, Wisniewski G, Rubin G, Jungreis CA, 
Wechsler LR, et al: Factors affecting survival rates for acute 
vertebrobasilar artery occlusions treated with intra-
artherial thrombolytic therapy: a meta-analytical approach. 
Neurosurgery 45:539–548, 1999

64. Liberati A, Altman DG, Tetzlaff J, Moher D, Tetzlaff J, Altman DG, 
et al: The PRISMA statement for reporting sys- 
tematic reviews and meta-analyses of studies that evaluate 
healthcare interventions: explanation and elaboration. BMJ 
339:b2700, 2009

65. Lucas CP, Zabramski JM, Spetzler RF, Jacobowitz R: Treatment 
of multiple systematic reviews (R-AMSTAR) for grad- 

FM, Moher D, et al: The medical review article revisited: has 
the medical review article revisited: clinical article. J 
Neurosurg 1994

67. McColgan P, Thant KZ, Sharma P: The genetics of sporadic 
ruptured and unruptured intracranial aneurysms: a genetic 
meta-analysis of 8 genes and 13 polymorphisms in approxi- 
ately 20,000 individuals. Clinical article. J Neurosurg 112: 
714–721, 2010

68. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup 
DF: Improving the quality of reports of meta-analyses of 
randomised controlled trials: the QUOROM statement. Lancet 
354:1896–1900, 1999

69. Moher D, Liberati A, Tetzlaff J, Altman DG: Preferred re- 
porting items for systematic reviews and meta-analyses: the 

70. Moher D, Schulz KF, Altman D: The CONSORT statement: 
revised recommendations for improving the quality of reports 
2001

71. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG: Epidemiology and reporting characteristics of systematic re- 

72. Mortazavi MM, Romeo AK, Deep A, Griessenuar CJ, Shoja 
MM, Tubbs RS, et al: Hypertonic saline for treating raised 
intracranial pressure: literature review with meta-analysis. A 

73. Noble AJ, Schenk T: Which studies help explain the poor 

versus nonoperative management of acute odontoid Type II 
11:651–658, 2009

75. Ooi YC, Dagi TF, Maltenfort M, Rincon F, Vibbert M, Jabbour 
P, et al: Tight glycemic control reduces infection and improves 
neurological outcome in critically ill neurosurgical and neuro-

76. Oxman AD, Guyatt GH: Validation of an index of the quality 

77. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison 
BG, Milner RA, et al: Agreement among reviewers of review 

78. Palur RS, Berk C, Schulzer M, Honey CR: A metaanalysis 
comparing the results of pallidotomy performed using mi-
icroelectrode recording or macroelectrode stimulation. J 
Neurosurg 96:1058–1062, 2002

79. Panyczkowski DM, Tomycz ND, Okonkwo DO: Comparative 
effectiveness of using computed tomography alone to exclude 
cervical spine injuries in obtunded or intubated patients: me-
ta-analysis of 14,327 patients with blunt trauma. A review. J 
Neurosurg 115:541–549, 2011

Gynecol 41:235–244, 1998

81. Quigley M: Risk of shunt-dependent hydrocephalus after oc- 
culsion of ruptured intracranial aneurysms by surgical clip-
ping or endovascular coiling: a single-institution series and 

82. Qureshi AI, Kirmani JF, Divani AA, Hobson RW II: Carotid 
angioplasty with or without stent placement versus carotid endarterectomy for treatment of carotid stenosis: a meta-

83. Rangoussi S, Li D, Webster AC: Perioperative epidural ste-
roids for lumbar spine surgery in degenerative spinal disease. A 

84. Ratilal B, Costa J, Sampaio C: Antibiotic prophylaxis for sur-
gical introduction of intracranial ventricular shunts: a system-

85. Ratliff JK, Cooper PR: Cervical laminoplasty: a critical re-

86. Ressing M, Blettner M, Klug SJ: Systematic literature reviews 
and meta-analyses: part 6 of a series on evaluation of scientific 

87. Rowland NC, Englot DJ, Mckinney PA, Sugarheue ME, Barbaro 
NM, Chang EF: A meta-analysis of predictors of seizure free-
dom in the surgical management of focal cortical dysplasia. 

88. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers 
TC: Meta-analyses of randomized controlled trials. N Engl J 

89. Sekula RF Jr, Bhatia S, Frederickson AM, Jannetta PJ, Quigley 
MR, Small GA, et al: Utility of intraoperative electromyogra-
phy in microvascular decompression for hemifacial spasm: a 
Methodology and reporting of neurosurgical meta-analyses


112. Whitehead A: Meta-Analysis of Controlled Clinical Trials. West Sussex: John Wiley & Sons, 2002


Manuscript submitted January 28, 2013. Accepted November 25, 2013. Please include this information when citing this paper: published online January 24, 2014; DOI: 10.3171/2013.11.JNS13195. Address correspondence to: Paul Klimo Jr., M.D., M.P.H., 6325 Humphreys Blvd., Memphis, TN 38120. email: pklimo@semmesmurphey.com.
AMSTAR: a measurement tool to assess the methodological quality of systematic reviews

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes/No Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was an “a priori” design provided?</td>
<td>The research question and inclusion criteria should be established before the conduct of the review. Note: Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a “yes.”</td>
</tr>
<tr>
<td>2. Was there independent study selection and data extraction?</td>
<td>There should be at least 2 independent data extractors and a consensus procedure for disagreements should be in place. Note: 2 people do study selection, 2 people do data extraction, consensus process or 1 person checks the other's work.</td>
</tr>
<tr>
<td>3. Was a comprehensive literature search performed?</td>
<td>At least 2 electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found. Note: If at least 2 sources + 1 supplementary strategy used, select “yes” (Cochrane register/Central counts as 2 sources; a gray literature search counts as supplementary).</td>
</tr>
<tr>
<td>4. Was the status of publication (i.e., gray literature) used as an inclusion criterion?</td>
<td>The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc. Note: If review indicates that there was a search for “gray literature” or “unpublished literature,” indicate ‘yes.” SIGLE database, dissertations, conference proceedings, and trial registries are all considered gray for this purpose. If searching a source that contains both gray and non-gray, must specify that they were searching for gray/unpublished lit.</td>
</tr>
<tr>
<td>5. Was a list of studies (included and excluded) provided?</td>
<td>A list of included and excluded studies should be provided. Note: Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select “no.”</td>
</tr>
<tr>
<td>6. Were the characteristics of the included studies provided?</td>
<td>In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed (e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases) should be reported. Note: Acceptable if not in table format as long as they are described as above.</td>
</tr>
<tr>
<td>7. Was the scientific quality of the included studies assessed and documented?</td>
<td>“A priori” methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant. Note: Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for each study (&quot;low&quot; or &quot;high&quot;) is fine, as long as it is clear which studies scored “low” and which scored &quot;high&quot;; a summary score/range for all studies is not acceptable.</td>
</tr>
<tr>
<td>8. Was the scientific quality of the included studies used appropriately in formulating conclusions?</td>
<td>The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations. Note: Might say something such as “the results should be interpreted with caution due to poor quality of included studies.” Cannot score “yes” for this question if scored “no” for question 7.</td>
</tr>
<tr>
<td>9. Were the methods used to combine the findings of studies appropriate?</td>
<td>For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., chi-squared test for homogeneity, I²). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?). Note: Indicate “yes” if they mention or describe heterogeneity, i.e., if they explain that they cannot pool because of heterogeneity/variability between interventions.</td>
</tr>
<tr>
<td>10. Was the likelihood of publication bias assessed?</td>
<td>An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges-Olkin). Note: If no test values or funnel plot included, score “no.” Score “yes” if mentions that publication bias could not be assessed because there were fewer than 10 included studies.</td>
</tr>
<tr>
<td>11. Was the conflict of interest included?</td>
<td>Potential sources of support should be clearly acknowledged in both the systematic review and the included studies. Note: To get a “yes,” must indicate source of funding or support for the systematic review and for each of the included studies.</td>
</tr>
</tbody>
</table>

# Methodology and reporting of neurosurgical meta-analyses

## Appendix 2

**PRISMA—checklist of items to include when reporting a systematic review or meta-analysis**

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item No.</th>
<th>Checklist Item</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td></td>
<td><strong>1</strong> Identify the report as a systematic review, meta-analysis, or both</td>
</tr>
<tr>
<td><strong>Abstract</strong></td>
<td></td>
<td><strong>2</strong> Provide a structured summary including, as applicable, background, objectives, data sources, study eligibility criteria, participants, interventions, study appraisal and synthesis methods, results, limitations, conclusions and implications of key findings, systematic review registration number</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td></td>
<td><strong>3</strong> Describe the rationale for the review in the context of what is already known</td>
</tr>
<tr>
<td><strong>Rationale</strong></td>
<td></td>
<td><strong>4</strong> Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td><strong>5</strong> Indicate if a review protocol exists, if and where it can be accessed (such as web address), and, if available, provide registration information including registration number</td>
</tr>
<tr>
<td><strong>Eligibility criteria</strong></td>
<td></td>
<td><strong>6</strong> Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale</td>
</tr>
<tr>
<td><strong>Information sources</strong></td>
<td></td>
<td><strong>7</strong> Describe all information sources (such as databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched</td>
</tr>
<tr>
<td><strong>Search</strong></td>
<td></td>
<td><strong>8</strong> Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated</td>
</tr>
<tr>
<td><strong>Study selection</strong></td>
<td></td>
<td><strong>9</strong> State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)</td>
</tr>
<tr>
<td><strong>Data collection process</strong></td>
<td></td>
<td><strong>10</strong> Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators</td>
</tr>
<tr>
<td><strong>Data items</strong></td>
<td></td>
<td><strong>11</strong> List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made</td>
</tr>
<tr>
<td><strong>Risk of bias in individual studies</strong></td>
<td></td>
<td><strong>12</strong> Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis</td>
</tr>
<tr>
<td><strong>Summary measures</strong></td>
<td></td>
<td><strong>13</strong> State the principal summary measures (such as risk ratio, difference in means).</td>
</tr>
<tr>
<td><strong>Synthesis of results</strong></td>
<td></td>
<td><strong>14</strong> Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I² statistic) for each meta-analysis</td>
</tr>
<tr>
<td><strong>Risk of bias across studies</strong></td>
<td></td>
<td><strong>15</strong> Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)</td>
</tr>
<tr>
<td><strong>Additional analyses</strong></td>
<td></td>
<td><strong>16</strong> Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td><strong>17</strong> Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram</td>
</tr>
<tr>
<td><strong>Study characteristics</strong></td>
<td></td>
<td><strong>18</strong> For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide the citations</td>
</tr>
<tr>
<td><strong>Risk of bias within studies</strong></td>
<td></td>
<td><strong>19</strong> Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12).</td>
</tr>
<tr>
<td><strong>Results of individual studies</strong></td>
<td></td>
<td><strong>20</strong> For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot</td>
</tr>
<tr>
<td><strong>Synthesis of results</strong></td>
<td></td>
<td><strong>21</strong> Present results of each meta-analysis done, including confidence intervals and measures of consistency</td>
</tr>
<tr>
<td><strong>Risk of bias across studies</strong></td>
<td></td>
<td><strong>22</strong> Present results of any assessment of risk of bias across studies (see item 15)</td>
</tr>
<tr>
<td><strong>Additional analysis</strong></td>
<td></td>
<td><strong>23</strong> Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression) (see item 16)</td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td></td>
<td><strong>24</strong> Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (such as health care providers, users, and policy makers)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item No.</th>
<th>Checklist Item</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discussion (continued)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>25</td>
<td>Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)</td>
</tr>
<tr>
<td>Conclusions</td>
<td>26</td>
<td>Provide a general interpretation of the results in the context of other evidence, and implications for future research</td>
</tr>
<tr>
<td><strong>Funding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>27</td>
<td>Describe sources of funding for the systematic review and other support (such as supply of data) and role of funders for the systematic review</td>
</tr>
</tbody>
</table>