

# Development of the Veritas plot and its application in cardiac surgery: an evidence-synthesis graphic tool for the clinician to assess multiple meta-analyses reporting on a common outcome

Sukhmeet S. Panesar,  
BSc (Hons.), MBBS\*†

Christopher Rao, BSc (Hons.),  
MBBS\*

Joshua A. Vecht, MRCS\*

Saqeb B. Mirza, MRCS, MSc‡

Gopalakrishnan Netuveli, MD, PhD§

Richard Morris, PhD¶

Joe Rosenthal, MSc, DRCOG, DFFP¶

Ara Darzi, KBE, HonFREng\*

Thanos Athanasiou, PhD\*

From the \*Department of Biosurgery and Surgical Technology, Imperial College London, the †Patient Safety Division, National Patient Safety Agency, the ‡Department of Orthopaedic Surgery, Royal National Orthopaedic Hospital NHS Trust, the §Department of Primary Care and Social Medicine, Imperial College London, and the ¶Department of Primary Care and Population Health Sciences, University College London, UK

Accepted for publication  
Jan. 29, 2009

## Correspondence to:

Dr. T. Athanasiou  
111 Gowan Ave.  
Fulham, London SW6 6RQ  
United Kingdom  
fax +44 207 886 6777  
t.athanasiou@imperial.ac.uk

**Background:** Meta-analyses may be prone to generating misleading results because of a paucity of experimental studies (especially in surgery); publication bias; and heterogeneity in study design, intervention and the patient population of included studies. When investigating a specific clinical or scientific question on which several relevant meta-analyses may have been published, value judgments must be applied to determine which analysis represents the most robust evidence. These value judgments should be specifically acknowledged. We designed the Veritas plot to explicitly explore important elements of quality and to facilitate decision-making by highlighting specific areas in which meta-analyses are found to be deficient. Furthermore, as a graphic tool, it may be more intuitive than when similar data are presented in a tabular or text format.

**Methods:** The Veritas plot is an adaption of the radar plot, a graphic tool for the description of multiattribute data. Key elements of meta-analytical quality such as heterogeneity, publication bias and study design are assessed. Existing qualitative methods such as the Assessment of Multiple Systematic Reviews (AMSTAR) tool have been incorporated in addition to important considerations when interpreting surgical meta-analyses such as the year of publication and population characteristics. To demonstrate the potential of the Veritas plot to inform clinical practice, we apply the Veritas plot to the meta-analytical literature comparing the incidence of 30-day stroke in off-pump coronary artery bypass surgery and conventional coronary artery bypass surgery.

**Results:** We demonstrate that a visually-stimulating and practical evidence-synthesis tool can direct the clinician and scientist to a particular meta-analytical study to inform clinical practice. The Veritas plot is also cumulative and allowed us to assess the quality of evidence over time.

**Conclusion:** We have presented a practical graphic application for scientists and clinicians to identify and interpret variability in meta-analyses. Although further validation of the Veritas plot is required, it may have the potential to contribute to the implementation of evidence-based practice.

**Contexte :** Les méta-analyses peuvent avoir tendance à produire des résultats trompeurs à cause de la rareté des études expérimentales (particulièrement en chirurgie), du biais de publication et de l'hétérogénéité de la conception des études, des interventions et des populations de patients visées par les études incluses. Lorsqu'on étudie une question clinique ou scientifique précise sur laquelle plusieurs méta-analyses pertinentes ont pu être publiées, il faut poser des jugements de valeur pour déterminer quelle analyse présente les données probantes les plus solides, et il faut reconnaître spécifiquement ces jugements de valeur. Nous avons conçu le schéma Veritas pour explorer implicitement les éléments importants de la qualité et faciliter la prise de décisions en mettant en évidence les aspects particuliers à l'égard desquels on constate des lacunes dans les méta-analyses. De plus, comme outil graphique, il sera peut-être plus intuitif que les tableaux ou textes qui présentent des données semblables.

**Méthodes :** Le schéma Veritas constitue une adaptation du schéma radar, outil graphique servant à distribuer des données à attributs multiples. On évalue les éléments clés de la qualité de la méta-analyse comme l'hétérogénéité, le biais de publication et la conception de l'étude. Des méthodes qualitatives existantes comme l'outil d'évaluation de multiples critiques systématiques (Assessment of Multiple Systematic Reviews ou AMSTAR) ont été intégrées, en plus de facteurs importants dans l'interprétation de méta-analyses en chirurgie, notamment l'année de publication et les caractéristiques démographiques. Afin de démontrer le potentiel qu'offre le schéma Veritas pour

éclairer la pratique clinique, nous l'appliquons aux méta-analyses qui comparent l'incidence de l'accident vasculaire cérébral à 30 jours dans les cas de pontage aortocoronarien sans circulation extracorporelle et de pontage aortocoronarien classique.

**Résultats :** Nous démontrons qu'un outil stimulant sur le plan visuel qui résume des données pratiques peut orienter le clinicien et le scientifique vers une méta-analyse particulièrement susceptible d'éclairer la pratique clinique. Le schéma Veritas est aussi cumulatif et nous a permis d'évaluer la qualité des données probantes au fil du temps.

**Conclusion :** Nous avons présenté une application graphique pratique à l'intention des scientifiques et des cliniciens qui leur permet de repérer et d'interpréter la variabilité dans les méta-analyses. Même si une validation plus poussée du schéma Veritas s'impose, elle pourrait contribuer à l'implantation de la pratique fondée sur les données probantes.

Identifying the highest quality evidence from the literature and applying it to clinical practice has always been a challenge for clinicians. The term "meta-analysis" was developed to describe the methodological framework for quantitative and systematic combination of results from previous research to reach conclusions about that body of research.<sup>1</sup> Meta-analysis progressed quickly and occupied an important place in surgical research. Proponents of meta-analysis argue that robust analyses of well conducted randomized controlled trials (RCTs) sit at the top of the "evidence hierarchy"<sup>2</sup> and should carry more weight than other study designs when determining the effectiveness of clinical interventions.<sup>3,4</sup>

### THE CHALLENGES OF APPLYING META-ANALYSIS IN SURGERY

The application of meta-analysis in surgery is challenging and can be particularly problematic. In the following sections we discuss important methodological controversies that need to be considered by clinicians and scientists seeking to implement evidence-based medicine to achieve improvements in the quality of health care delivery and clinical care.

#### *The "garbage in-garbage out" effect*

The dependence of meta-analysis on the quality and characteristics of included studies was famously described as "garbage in-garbage out."<sup>5</sup> The application of meta-analysis to nonexperimental studies in particular is controversial, as the results of nonexperimental studies are vulnerable to bias by confounding factors.<sup>6</sup> This is particularly important in surgical meta-analysis; the nature of surgical interventions often make it difficult to perform well-conducted RCTs, as it would be unethical to randomly assign patients to groups in which they may be subject to potentially harmful risks.<sup>7</sup> This is reflected in the relative paucity of surgical experimental studies compared with medical studies.<sup>8,9</sup> Most literature assessing the efficacy of surgical interventions consists of retrospective case series, with RCTs accounting for less than 10% of the total.<sup>8,10-12</sup> Whereas there are guidelines to assist with pre-

paration of both randomized (QUOROM statement)<sup>13</sup> and observational studies (MOOSE statement),<sup>14</sup> these are primarily targeted at researchers conducting meta-analyses and do not facilitate rapid assessment of the quality and type of included studies.

#### *Heterogeneity or the "apples and oranges" effect*

Heterogeneity in study design, intervention or patient population can result in the comparison of "apples and oranges" and may substantially affect the conduct and results of a meta-analysis.<sup>15</sup> The perception of statistical heterogeneity often influences meta-analysts and clinicians in important decisions. These decisions include whether data extracted from different studies are similar enough to combine and whether the treatment advocated is applicable to the general targeted population.<sup>16</sup>

The recognition of heterogeneity can also be considered a positive finding. It widens the spectrum of the meta-analytical results. It may also highlight factors that influence outcome that were not observable in individual trials. Furthermore, if the effect is consistent even with discrepant studies, it strengthens the case for the causality of the treatment. Finally, if a meta-analysis is performed before beginning a new study, then heterogeneity may help the investigator improve his design by incorporating an understanding of these other factors.<sup>17</sup> A commonly used test for identifying the degree of heterogeneity is the  $I^2$  test; however, this is often not calculated.<sup>18</sup>

#### *Publication bias and the "file drawer problem"*

The exclusion of studies from meta-analyses can bias the results. This can occur for 2 reasons: first, because of a flawed, incomplete literature search and second, because of publication bias or the "file drawer problem."<sup>19</sup> Publication bias is an inherent issue in meta-analyses. Studies are often not published in indexed journals and are consequently not included in meta-analyses if they do not report original, positive or statistically significant findings. If publication bias occurs, the subsequent meta-analysis of published literature may be misleading as it

may not accurately reflect the available evidence.<sup>20</sup>

The simplest and most commonly used method to detect publication bias is an informal examination of a funnel plot. Formal tests for publication bias exist, but in practice few meta-analyses have assessed or adjusted for the presence of publication bias. A recent assessment of the quality of systematic reviews stated that only 6.5% of studies in high-impact general journals and 3.2% in specialty journals reported that a funnel plot had been examined.<sup>21</sup>

Variations in individual study characteristics may substantially affect one or all of these previously mentioned primary dimensions of meta-analytical quality. Secondary meta-analytical controversies include the chronology of meta-analyses and the patient population included. Often these factors are not considered in sufficient depth and may result in poor inferences being drawn from meta-analyses.

### *Study population*

The contribution of the study population in individual studies on the results of meta-analyses is important as it is a major determinant of heterogeneity. For example, if the targeted intervention is only beneficial in high-risk groups such as the elderly,<sup>22</sup> it would be hazardous to apply results of such a meta-analysis to low-risk groups. Furthermore, greatly differing event rates among populations reported in different studies can potentially bias the meta-analysis.<sup>23,24</sup> This is particularly important in surgical meta-analysis because event rates can vary substantially depending on patient comorbidities.

### *Year of publication*

The chronology of meta-analysis is important for a number of reasons. The year a study is published is a significant determinant of heterogeneity as population characteristics may change over time, whereas the development in technology and technical expertise may translate into unfavourable outcomes in early studies. These factors need to be taken into account, especially in surgical specialties where new technologies and techniques are continuously developed and the learning curve is progressively overcome.<sup>25</sup> Finally, as the pool of evidence accumulates with time, the summary results reported by meta-analyses will be refined.<sup>26</sup>

Meta-analysis is further complicated, as attempts to adhere to all dimensions of primary and secondary meta-analytical quality will often suggest contradictory study inclusion and exclusion criteria and encourage different meta-analytical conduct. Consequently, meta-analysis, especially surgical meta-analysis, is often an exercise in compromise. For example, inclusion of only experimental trials may exclude the most contemporary literature, or an attempt to include all available evidence may introduce heterogeneity. Qualitative value judgments interpreting

the relative importance of dimensions of quality are required even in the structure of a statistically robust meta-analysis that integrates a comprehensive literature search. Therefore a compromise is necessary for qualitative value judgments. This does not necessarily demean the effectiveness of meta-analysis as a tool for literature review, evidence synthesis and clinical decision-making; however, it is imperative that these value judgments are explained and justified. Unfortunately, this is often not the case.<sup>18,21</sup>

Similarly, when investigating a specific clinical question on which several relevant meta-analyses may have been published, value judgments must be applied to determine which of the meta-analyses represents the most robust evidence. These value judgments should also be explicitly acknowledged.

## **QUALITY ASSESSMENT TOOLS OF META-ANALYSES**

More than 20 tools are available to assess the quality of systematic reviews.<sup>27</sup> Nevertheless, most of the available instruments are not widely used. Several are lengthy and include complicated instructions. One of the most commonly used is the Overview Quality Assessment Questionnaire (OQAQ).<sup>28</sup> An improved version of this tool is the recently developed and validated Assessment of Multiple Systematic Reviews (AMSTAR) tool,<sup>29</sup> an 11-item instrument that assesses key attributes of a well-conducted meta-analysis. It assesses whether an a priori design was provided before conduction of the meta-analysis, whether there was any duplicate selection or data extraction and whether a comprehensive literature search was used. Furthermore, it asks whether the authors made an attempt to source grey literature. It looks at whether a list of included and excluded studies was provided and whether characteristics of the included studies were provided. Also, it determines whether the quality of the included studies was assessed, documented and used appropriately in formulating questions. Finally, it evaluates whether the methods used to combine the findings of the studies were appropriate, whether publication bias was assessed and conflict of interest stated. Each question has 4 responses: “yes,” “no,” “can’t answer” and “not applicable.” A “yes” gives a score of 1; any other response results in a score of 0. The overall score is out of 11.<sup>29</sup> The AMSTAR tool too has its limitations, however, which include minimal use to date, lack of graphical output and failure to fully assess important dimensions of quality (e.g., how the study deals with heterogeneity). Furthermore, it is conceivable that in different clinical specialties the relative importance of dimensions of meta-analytical quality may differ. For example, in a rapidly evolving specialty, it may be considered more important to include the most contemporary literature, whereas if several well-conducted randomized studies exist the quality of included studies may be considered more important. The AMSTAR tool is presented in Appendix 1.

**THE VERITAS PLOT**

We designed the Veritas plot to address the limitations of previous scoring systems, explicitly explore important elements of quality and facilitate decision-making by highlighting specific areas in which meta-analyses addressing similar clinical questions are found to be deficient. Furthermore, as a graphic tool we hoped that the Veritas plot would be more intuitive than presenting similar data in a tabular or text format.

In this paper, we demonstrate how the Veritas plot can be applied to an important example from the cardiothoracic literature. In addition, we discuss the strengths, limitations and potential for further development of the Veritas plot.

**METHODS**

In this section we illustrate the key methodological steps required to create a Veritas plot, using an example from the cardiothoracic literature. By comparing the quality of meta-analyses that investigate the effect of off-pump coronary bypass techniques on the incidence of stroke, we hope to demonstrate the role of the Veritas plot in clinical decision-making.

The Veritas plot can be considered to be a variant of the radar or radial plot. A radar plot is a graphic display for comparing estimates that have differing precisions.<sup>30</sup> Radar plots have existed for many years and are an important descriptive tool for multiattribute data. In general, the common feature of radar plots is that they are a circular graphing method and have a series of spokes or rays projecting from a central point, with each ray representing a different variable label. The values of the variables are encoded into the lengths of the rays, and the values so plotted are sometimes connected to form an enclosed figure.<sup>31</sup> In the Veritas plot, the rays represent dimensions of meta-analytical quality.

In our example, we included the 6 dimensions of meta-analytical quality that most influence the final outcome. As important elements of meta-analytical quality, we included assessments of whether the studies consider publication bias and heterogeneity, and we have considered the design of studies included in the analysis. We included the AMSTAR score as one of our dimensions of quality, as it is a strong tool for assessing the quality and comprehensiveness of the literature review. Finally, we included the baseline population event risk and year of publication as further dimensions of meta-analytical quality, as meta-analyses addressing the effect of rapidly evolving technology on infrequently occurring side-effects could easily be biased by these factors. Our methods included the following steps.

**1. Clinical question**

A clinical question is raised (e.g., what is the difference in

outcome of 30-day stroke between cardiac surgery patients undergoing conventional coronary artery bypass (CCAB) and off-pump coronary artery bypass (OPCAB) surgery?).

**2. Search strategy**

A search strategy must be conducted to investigate the clinical question. In our example, we used an expert search strategy as described by Kelly and colleagues<sup>32</sup> to search for any relevant meta-analyses published up to and including March 2008. We searched the following databases: MEDLINE, EMBASE, CINAHL and Google Scholar. We searched for meta-analyses either specifically examining the outcome of 30-day stroke or several outcomes, including 30-day stroke. We limited the search to include “meta-analysis,” “human only studies” and “English language.” The results of our search are shown in Figure 1.

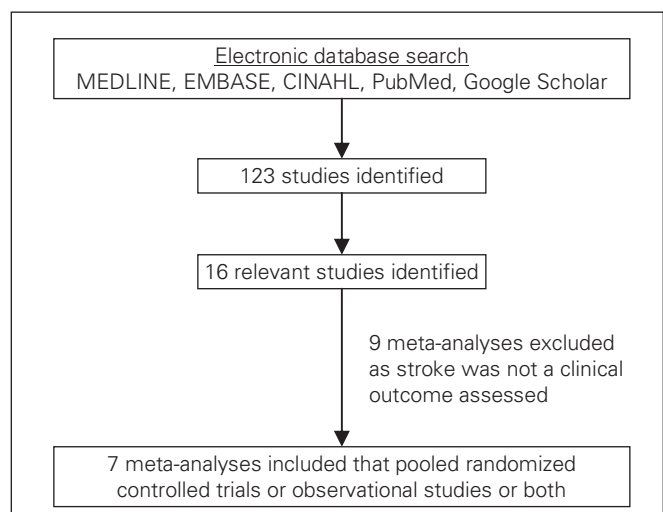
**3. Ranking of studies**

The studies identified require ranking. In our example, we identified 7 studies and ranked them according to 6 categories: study type, publication bias, heterogeneity, year of publication, risk of population included and AMSTAR score. The characteristics of the studies included are presented in Table 1.

The scoring system worked as follows. When analyzing each category, the study with the best score received  $n$  points where  $n$  = the number of studies. The second best study received  $n - 1$  points, and so on. In the case of 2 studies performing equally well, the study with the next highest score would receive  $n - 2$  points.

**4. Implementation of data set in a radar plot**

We constructed the radar plots using Microsoft Excel software (Microsoft Corp.).



**Fig. 1.** Study selection results.



### 5. Formatting the Veritas plot

We created the Veritas plot as a variant of the radial plot. Individual Veritas plots for all of the studies are shown in Figure 2. We showed the score of a study in each dimension of quality with a solid line, which can be compared with the triangles that represent the mean score across all 7 studies in each of the dimensions of quality. This way individual study performance in each dimension of quality can be evaluated.

We conceived the Veritas score as a summary statistic. It is calculated by taking the mean of the scores in the 6 dimensions of quality for each study. It is represented in the plots shown in Figure 2 by dashed lines. It can be compared with the mean Veritas score across all 7 studies (represented by the squares in Fig. 2) to give an overall estimate of the study quality. In Figure 3, the Veritas scores for all of the studies are plotted. For ease of interpretation,

the scores of the individual studies in each of the dimensions of quality were not plotted. This figure can be used to compare overall quality in the included studies across all 6 selected dimensions of quality.

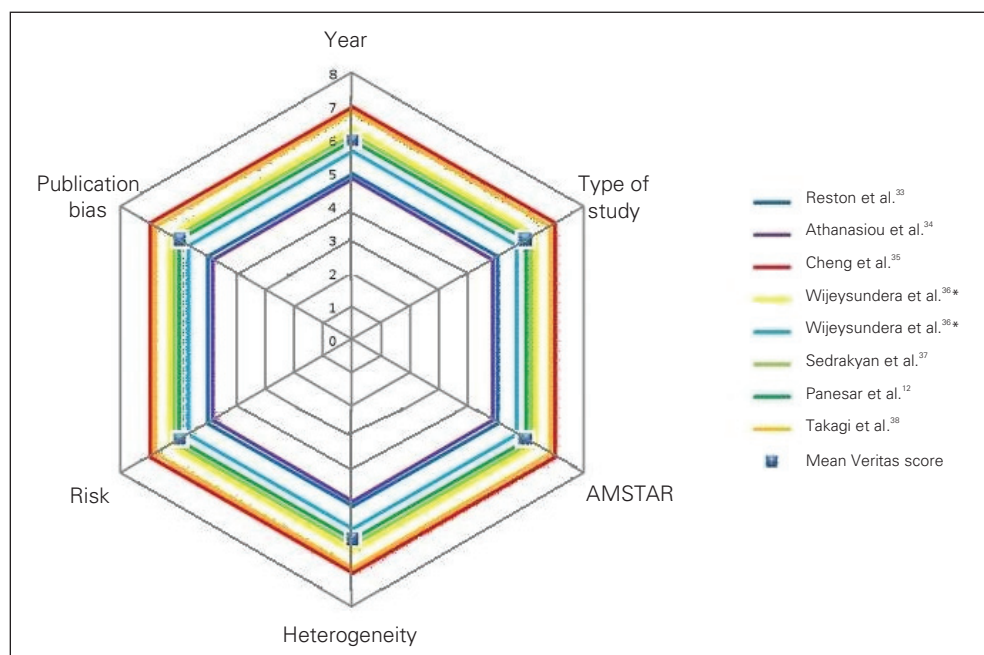
### 6. Interpretation of the Veritas plot

The greater the score of a study in a specific quality dimension, and hence the further the solid plot is from the central pole, the better its performance in this quality dimension. If the solid plot lies outside of the triangles, which represent the mean score across all 7 studies in each dimension, then it could be said to perform “better than average” in this dimension. The Veritas score is shown with a dashed plot. Similarly, if it lies outside of the squares, which represent the mean Veritas score across all 7 studies, then the study could be said to be “better than

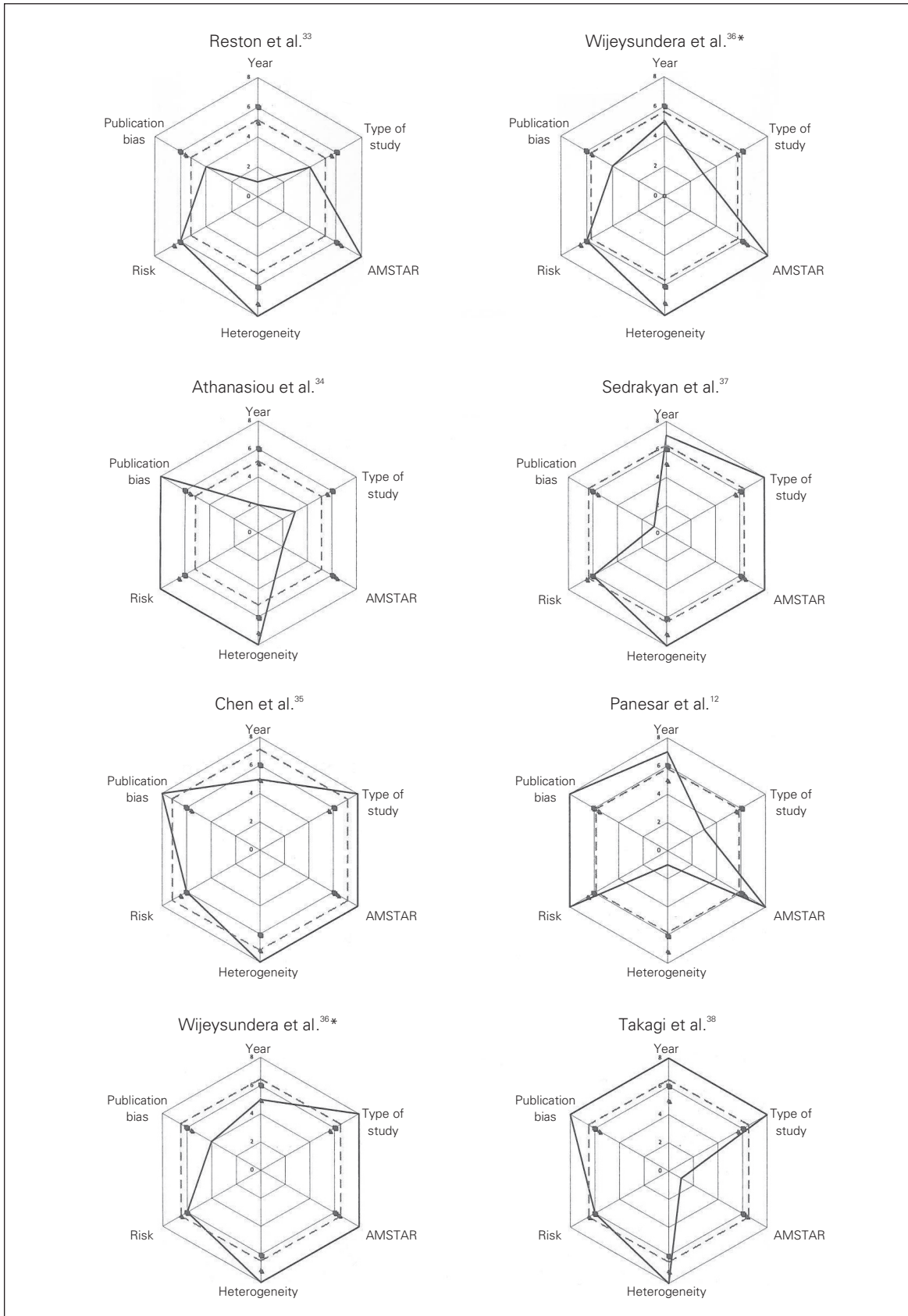
**Table 1. Characteristics of studies included in our example**

Study	Study characteristic (score)										Veritas score
	Year of publication	Type of study	AMSTAR score	Heterogeneity	Population risk	Publication bias					
Reston et al. <sup>33</sup>	2003	(1) Mixed	(4)	8 (8)	Homogenous (8)	Low (6)	Partially assessed (4)				5.17
Athanasίου et al. <sup>34</sup>	2004	(2) Observational	(3)	7 (2)	Homogenous (8)	High (8)	Assessed (8)				5.17
Cheng et al. <sup>35</sup>	2005	(5) Randomized	(8)	8 (8)	Homogenous (8)	Low (6)	Assessed (8)				7.17
Wijeyesundera et al. <sup>36*</sup>	2005	(5) Randomized	(8)	8 (8)	Homogenous (8)	Low (6)	Partially assessed (4)				6.50
Wijeyesundera et al. <sup>36*</sup>	2005	(5) Observational	(3)	8 (8)	Homogenous (8)	Low (6)	Partially assessed (4)				5.67
Sedrakyan et al. <sup>37</sup>	2006	(7) Randomized	(8)	8 (8)	Homogenous (8)	Low (6)	Not assessed (1)				6.33
Panesar et al. <sup>22</sup>	2006	(7) Observational	(3)	8 (8)	Heterogeneous (1)	High (8)	Assessed (8)				5.83
Takagi et al. <sup>38</sup>	2007	(8) Randomized	(8)	5 (1)	Homogenous (8)	Low (6)	Assessed (8)				6.50

AMSTAR = the Assessment of Multiple Systematic Reviews tool.<sup>29</sup>  
 \*This study considered randomized controlled trials and observational studies separately.



**Fig. 2.** Veritas plots for included studies. \*This study considered randomized controlled trials and observational studies separately. AMSTAR = Assessment of Multiple Systematic Reviews tool.



**Fig. 3.** Comparison of Veritas score in all included studies.\*This study considered randomized controlled trials and observational studies separately. AMSTAR = Assessment of Multiple Systematic Reviews tool.

average.” It is important to note that although the plot is effective in demonstrating the dimensions of quality in which the study performs well, the Veritas score has not been validated as a summary statistic. The Veritas plot is intended to facilitate value judgment in the quality assessment of meta-analysis to inform clinical practice and is not intended to replace such judgments. Consequently the Veritas score must be interpreted with caution.

## RESULTS

The results of our search strategy are shown in Figure 1. We identified 16 meta-analyses comparing OPCAB with CCAB surgery. Of these, 7 meta-analyses studied stroke.<sup>22,33–38</sup> One of the studies considered RCTs and observational studies separately.<sup>36</sup>

We developed individual Veritas plots for each of the 7 studies (Fig. 2). The meta-analysis by Cheng and colleagues<sup>35</sup> scored the highest in 4 of the 6 dimensions and had the highest Veritas score: 7.17.

## DISCUSSION

The advantages of well-conducted meta-analyses are that they allow objective appraisal of the evidence in comparison with traditional narrative reviews, provide a more precise estimate of a treatment effect and may explain heterogeneity among the results of individual studies. The application of meta-analysis in surgery, however, is problematic, and surgical meta-analyses are often vulnerable to bias. Patient population characteristics, technological development and evolving surgical expertise have the potential to substantially affect event rates, which can bias the results of meta-analyses.<sup>23,24</sup>

There are numerous instances where meta-analyses have pooled results from small trials with disparate results and, as such, have produced conflicting evidence.<sup>39–43</sup> Furthermore, results have been generated that conflicted with those of subsequent large RCTs.<sup>44–47</sup> When this occurs, the reliability of the evidence is questioned, resulting in poorly guided clinical decisions. Consequently, doubts have been raised about the reliability of using meta-analyses to guide clinical practice.<sup>48,49</sup> Although even advocates of meta-analysis argue that such studies are not a substitute for RCTs,<sup>50</sup> meta-analyses may be a useful guide to clinical decision-makers until unequivocal experimental evidence is available.<sup>6</sup> However, if meta-analysis is to continue to have a role in surgical decision-making, surgeons need to be able to assess, compare and communicate the quality of meta-analyses, particularly in areas where several meta-analyses are available.

### *Strengths of the Veritas plots*

We have presented a novel application of radar plots that clinicians can use when they are presented with results

from several meta-analyses. Our Veritas plot enables the clinician to assess key attributes that evaluate the quality of a meta-analysis. This tool helps with the analytical process and allows clinicians to decide on the translation of results from the meta-analyses to patient care.

In addition to text or tables, research results can also be presented in graphic formats. Graphic displays are particularly suitable for illustrating relations and trends concisely.

Our tool has several advantages. First, it incorporates the different facets that constitute a well-conducted meta-analysis without the need for extensive narrative. It considers important elements of quality such as heterogeneity, publication bias, quality scoring, year of publication and the risk stratification of the group. It builds on the strengths of current quality-assessment tools such as the AMSTAR.<sup>27–29</sup> It is an invaluable tool for appraising the ever-increasing number of meta-analyses in fields such as OPCAB surgery. Hence, one can say that it has a cumulative potential. It is also a versatile tool and in meta-analyses where attributes such as year of publication are not important, these can be removed. We believe that it is also a quick and simple tool that can be applied across any surgical discipline. Further benefit occurs by virtue of its graphic representation; our method makes it easier for the clinician to appraise the quality of meta-analyses.<sup>51</sup>

We believe that this tool builds on the foundations of the evidence-based medicine hierarchy and allows for effective communication between researchers and clinicians.<sup>52</sup> It also allows interstudy variation to be assessed effectively and the findings of the study to be applied accordingly to daily surgical practice.

### *Limitations of the Veritas plots*

Further validation of the Veritas plot by appraisers of meta-analyses is needed to assess its validity, reliability and perceived utility. Its application needs to be verified across specialties, and its user-friendliness should be confirmed. Furthermore, this tool is subjective; authors in other fields may dispute the emphasis we have placed on the attributes in our Veritas plot. The issue of scaling and appropriate weighting of attributes will need further validation.

At present this tool cannot compare the quality of studies from different clinical areas. Of equal importance is the fact that it relies on a study being ranked relative to others. Consequently, all dimensions of quality need to be assessed and obtained.

## CONCLUSION

The Veritas plot is a practical, novel and useful aid in the quality assessment of several meta-analyses studying the same outcome. It was designed to consider important elements of meta-analytical quality such as heterogeneity, publication bias, study design, quality scoring, year of

publication and population characteristics. The method is suitable, however, for adaptation to a variety of questions in evidence synthesis.

We therefore invite colleagues to consider applying and adapting the Veritas plot as a component of the processes of synthesizing and reporting the findings of multiple meta-analyses assessing similar outcomes.

**Competing interests:** None declared.

**Contributors:** Dr. Panesar was responsible for fundamental conceptual development, intellectual content, statistical analysis, data collection and interpretation and manuscript drafting. Dr. Rao was responsible for fundamental conceptual development, intellectual content, statistical analysis, data interpretation and manuscript and figure drafting. Drs. Vecht, Mirza, Netuveli, Morris and Rosenthal were responsible for data collection and preliminary analysis and provided important intellectual content. Drs. Netuveli and Morris are statisticians. Lord Darzi was responsible for providing important intellectual content throughout the manuscript's production and for final approval of the version to be published. The Veritas Plot was conceived by Dr. Athanasiou, whose involvement was critical to every phase of this work. All authors read and approved the final manuscript.

References

1. Glass G. Primary, secondary and meta-analysis of research. *Educ Res* 1976;5:3-8.
2. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995;274:1800-4.
3. Jones RS, Richards R. Office of evidence based surgery charts course for improved system of care. *Bull Am Coll Surg* 2003;88:11-21.
4. Sackett DL. Applying overviews and meta-analyses at the bedside. *J Clin Epidemiol* 1995;48:61-6.
5. Eysenck H. An exercise in mega-silliness. *Am Psychol* 1978;33:517.
6. Petitti D. *Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine*. Oxford (UK): Oxford University Press; 2000.
7. Berman NG, Parker RA. Meta-analysis: neither quick nor easy. *BMC Med Res Methodol* 2002;2:10.
8. Ellis J, Mulligan I, Rowe J, et al. Inpatient general medicine is evidence based. A-Team, Nuffield Department of Clinical Medicine. *Lancet* 1995;346:407-10.
9. Howes N, Chagla L, Thorpe M, et al. Surgical practice is evidence based. *Br J Surg* 1997;84:1220-3.
10. Pollock AV. Surgical evaluation at the crossroads. *Br J Surg* 1993; 80:964-6.
11. Pollock AV. The rise and fall of the randomized control trial in surgery. *Theor Surg* 1989;4:163-70.
12. Panesar SS, Thakrar R, Athanasiou T, et al. Comparison of reports of randomized controlled trials and systematic reviews in surgical journals: literature review. *J R Soc Med* 2006;99:470-2.
13. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999;354:1896-900.
14. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.
15. Wachter KW. Disturbed by meta-analysis? *Science* 1988;241:1407-8.
16. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* 2007;335:914-6.
17. Deeks JJ, Higgins JPT, Altman DG, editors. 9.5.2. Identifying and

- measuring heterogeneity. In: Higgins JPT, Green S, editors. *The Cochrane handbook for systematic reviews of interventions*. Version 5.0.0 (updated February 2008). The Cochrane Collaboration, 2008. Most recent update available from www.cochrane-handbook.org.
18. Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 1991;338:1127-30.
19. Rao C. Selection models and the file drawer problem [comment]. *Statist Sci* 1988;3:131. Comment on: *Statist Sci* 1988;1:109-17.
20. Peters JL, Sutton AJ, Jones DR, et al. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006;295:676-80.
21. Sutton AJ, Duval SJ, Tweedie RL, et al. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;320:1574-7.
22. Panesar SS, Athanasiou T, Nair S, et al. Early outcomes in the elderly: a meta-analysis of 4921 patients undergoing coronary artery bypass grafting—comparison between off-pump and on-pump techniques. *Heart* 2006;92:1808-16.
23. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: How do their results compare? *JAMA* 1996;276:1332-8.
24. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. Examples from the management of acute myocardial infarction. *Ann Intern Med* 1995;123:873-7.
25. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: One answer is not always enough. *Lancet* 1998;351:123-7.
26. Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; 327:248-54.
27. Shea B, Moher D. *Assessing the quality of reports of systematic reviews: the QUORUM statement compared to other tools*. London (UK): BMJ Books; 2001.
28. Oxman AD. Checklists for review articles. *BMJ* 1994;309:648-51.
29. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
30. Galbraith R. Some applications of radial plots. *J Am Stat Assoc* 1994; 89:1232-4.
31. Saary MJ. Radar plots: a useful way for presenting multivariate health care data. *J Clin Epidemiol* 2008;61:311-7.
32. Kelly KD, Travers A, Dorgan M, et al. Evaluating the quality of systematic reviews in the emergency medicine literature. *Ann Emerg Med* 2001;38:518-26.
33. Reston JT, Tregear SJ, Turkelson CM. Meta-analysis of short-term and mid-term outcomes following off-pump coronary artery bypass grafting. *Ann Thorac Surg* 2003;76:1510-5.
34. Athanasiou T, Al Ruzzeq S, Kumar P, et al. Off-pump myocardial revascularization is associated with less incidence of stroke in elderly patients. *Ann Thorac Surg* 2004;77:745-53.
35. Cheng DC, Bainbridge D, Martin JE, et al. Does off-pump coronary artery bypass reduce mortality, morbidity, and resource utilization when compared with conventional coronary artery bypass? A meta-analysis of randomized trials. *Anesthesiology* 2005;102:188-203.
36. Wijesundera DN, Beattie WS, Djaiani G, et al. Off-pump coronary artery surgery for reducing mortality and morbidity: meta-analysis of randomized and observational studies. *J Am Coll Cardiol* 2005;46:872-82.
37. Sedrakyan A, Wu AW, Parashar A, et al. Off-pump surgery is associated with reduced occurrence of stroke and other morbidity as compared with traditional coronary artery bypass grafting: a meta-analysis of systematically reviewed trials. *Stroke* 2006;37:2759-69.
38. Takagi H, Tanabashi T, Kawai N, et al. Off-pump surgery does not reduce stroke, compared with results of on-pump coronary artery bypass grafting: a meta-analysis of randomized clinical trials. *J Thorac Cardiovasc Surg* 2007;134:1059-60.
39. McIntyre LA, Fergusson DA, Hebert PC, et al. Prolonged therapeutic hypothermia after traumatic brain injury in adults: a systematic review. *JAMA* 2003;289:2992-9.
40. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant



- systematic reviews. *CMAJ* 1997;156:1411-6.
41. Henderson WR, Dhingra VK, Chittock DR, et al. Hypothermia in the management of traumatic brain injury. A systematic review and meta-analysis. *Intensive Care Med* 2003;29:1637-44.
  42. Christensen E. Quality of reporting of meta-analyses: the QUOROM statement. Will it help? *J Hepatol* 2001;34:342-5.
  43. Delaney A, Bagshaw SM, Ferland A, et al. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care* 2005;9:R575-82.
  44. Egger M, Davey SG, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
  45. Furukawa TA, Streiner DL, Hori S. Discrepancies among meta-analyses. *J Clin Epidemiol* 2000;53:1193-9.
  46. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995;345:772-6.
  47. LeLorier J, Gregoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-42.
  48. Egger M, Smith GD. Meta-analysis. Potentials and promise. *BMJ* 1997;315:1371-4.
  49. Bailar JC III. The promise and problems of meta-analysis. *N Engl J Med* 1997;337:559-61.
  50. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995;48:23-40.
  51. Puhan MA, ter Riet G, Eichler K, et al. More medical journals should inform their contributors about three key principles of graph construction. *J Clin Epidemiol* 2006;59:1017-22.
  52. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. 1996. *Clin Orthop Relat Res* 2007; 455:3-5.

### Appendix 1. The Assessment of Multiple Systematic Reviews (AMSTAR) tool

#### 1. Was an "a priori" design provided?

The research question and inclusion criteria should be established before the conduct of the review.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 2. Was there duplicate study selection and data extraction?

There should be at least 2 independent data extractors, and a consensus procedure for disagreements should be in place.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 3. Was a comprehensive literature search performed?

At least 2 electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE and MEDLINE). Keywords 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.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 4. Was the status of publication (i.e., grey literature) used as an inclusion criterion?

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.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 5. Was a list of studies (included and excluded) provided?

A list of included and excluded studies should be provided.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 6. Were the characteristics of the included studies provided?

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.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 7. Was the scientific quality of the included studies assessed and documented?

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.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 8. Was the scientific quality of the included studies used appropriately in formulating conclusions?

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.

- Yes  
 No  
 Can't answer  
 Not applicable

#### 9. Were the methods used to combine the findings of studies appropriate?

For the pooled results, a test should be done to ensure the studies were combinable to assess their homogeneity (i.e.,  $\chi^2$  test for homogeneity,  $I^2$ ). 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?).

- Yes  
 No  
 Can't answer  
 Not applicable

#### 10. Was the likelihood of publication bias assessed?

An assessment of publication bias should include a combination of graphic aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

- Yes  
 No  
 Can't answer  
 Not applicable

#### 11. Was the conflict of interest stated?

Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

- Yes  
 No  
 Can't answer  
 Not applicable