Systematic Reviews Explained: AMSTAR—How to Tell the Good From the Bad and the Ugly

Mohammad O Sharif[1], Fyeza N Janjua Sharif[2], Hesham Ali[3], Farooq Ahmed[4]

[1]MSc, BDS (Hons), MJDF. National Institute for Health Research Academic Clinical Fellow—Orthodontics, School of Dentistry, University of Manchester, Manchester, UK. [2]BDS. General Dental Practitioner, Manchester, UK. [3]BDS, MFDS. Dental Foundation Year Two Trainee, Royal Manchester Children's Hospital, UK. [4]BDS(Hons), MFDS. Orthodontic Specialty Registrar, Manchester, UK.

Summary

Systematic reviews are essential in summarising evidence and providing an indication of its strength and direction. This is why they often inform clinical decision making. Although the quantity of reviews published is increasing, concerns about their quality may sometimes be questioned. This paper highlights the aspects of systematic review methodology that influence a review's overall quality. The authors explain the recently developed tool "Assessment of Multiple Systematic Reviews" (AMSTAR) to demonstrate how this can be used efficiently, allowing a busy clinician to evaluate quality and decide whether or not a particular review should be used to inform their clinical practice. Systematic reviews may allow clinicians to incorporate the best available evidence into clinical practice. The ability to evaluate the quality and reliability of systematic reviews is imperative in this process. The authors have used items detailed in AMSTAR to demonstrate the aspects of systematic review methodology that influence the overall quality of a review.

Key Words: Systematic Review, Quality, AMSTAR, Methodology

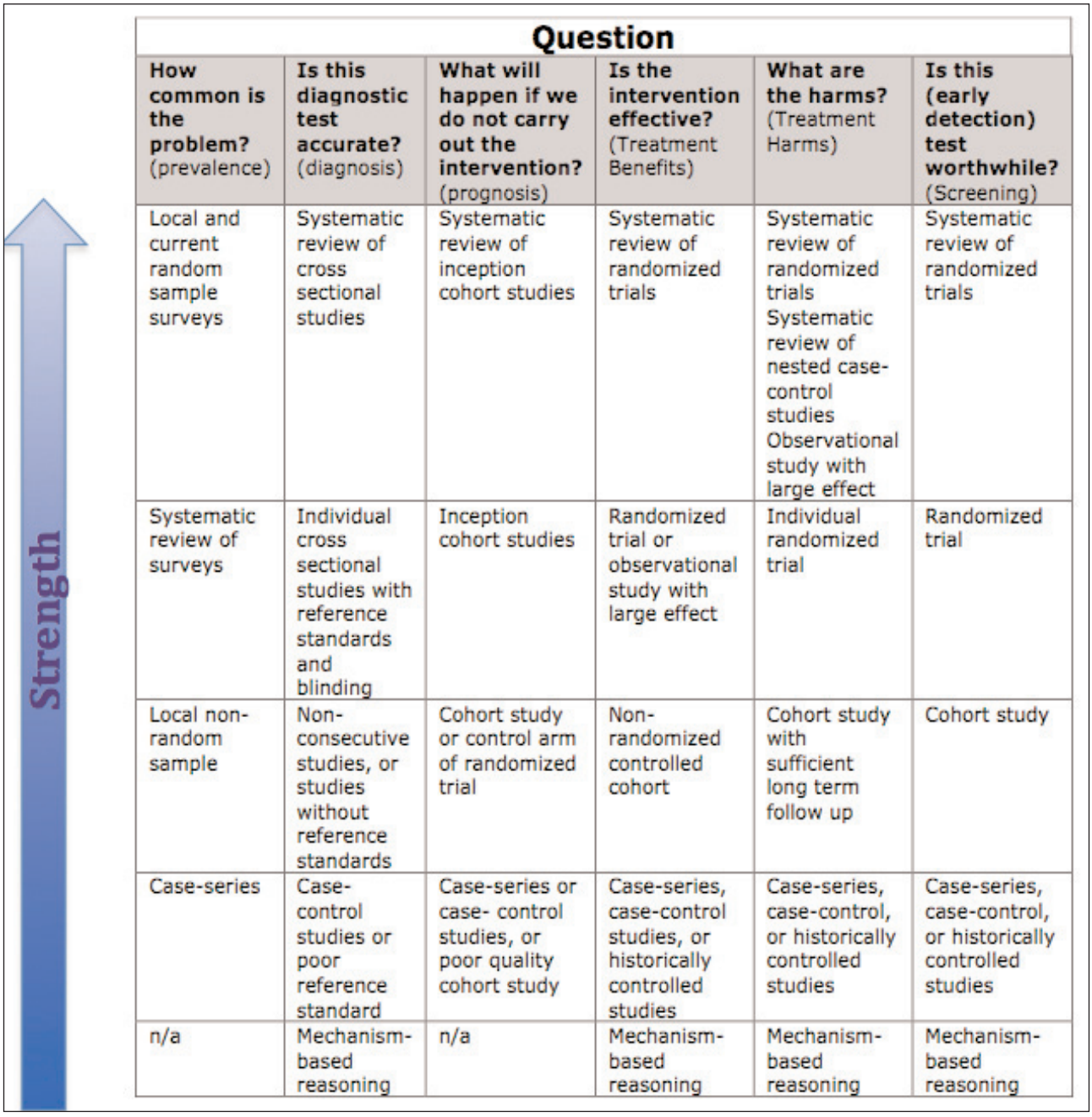
Introduction

Annually, 2,500 systematic reviews are added to the National Library of Medicine's PubMed MEDLINE (in English) database [1]. Systematic reviews can be invaluable for evaluating available evidence in a methodical manner and providing a critical summary of strength and direction of evidence [2]. Systematic reviews primarily evaluate the effects of an intervention for the prevention, treatment and rehabilitation of a condition. However, they can also be used to assess the accuracy of diagnostic tests, prognosis of a condition and aetiology [1].

The hierarchy of evidence varies depending on the nature of the question to be investigated (*Figure 1*). For interventional studies systematic reviews of randomised controlled trials are at the top of the hierarchy of evidence [2]. They are therefore regarded as the best source of evidence. However, this position is based on the presumption that they have been designed and conducted to the highest standards. Unfortunately, as with any form of evidence, systematic reviews vary in their quality and subsequently in their value (for guiding decision making). When considering systematic reviews, the Cochrane Collaboration systematic reviews [3] are considered to be of the highest quality (often termed "the gold standard") because they are conducted using set guidelines [4,5] and are independently peer-reviewed and published at both the protocol and completion phase. This process helps to ensure adherence to the criteria [5].

The reality is that the majority of published systematic reviews are non-Cochrane reviews and so it is important for a busy clinician to acknowledge that some of these reviews may fall short in their methodological standards. Because of this, such reviews may present a distorted view of the evidence underpinning a subject and hence draw inappropriate conclusions [6]. In fact, the term "Garbage in Garbage Out" (GIGO) has been used to illustrate the problem associated with poorly designed systematic reviews. In 2007, this concept was reported in a paper that highlighted concerns about systematic reviews in the field of endodontics; the author identified examples of "biased sam-

Corresponding author: Mohammad Owaise Sharif, National Institute for Health Research, Academic Clinical Fellow (NIHR ACF)—Orthodontics, The University of Manchester, Manchester M13 9PL, UK; e-mail: mohammad.owaise.sharif@googlemail.com



Question					
How common is the problem? (prevalence)	**Is this diagnostic test accurate?** (diagnosis)	**What will happen if we do not carry out the intervention?** (prognosis)	**Is the intervention effective?** (Treatment Benefits)	**What are the harms?** (Treatment Harms)	**Is this (early detection) test worthwhile?** (Screening)
Local and current random sample surveys	Systematic review of cross sectional studies	Systematic review of inception cohort studies	Systematic review of randomized trials	Systematic review of randomized trials Systematic review of nested case-control studies Observational study with large effect	Systematic review of randomized trials
Systematic review of surveys	Individual cross sectional studies with reference standards and blinding	Inception cohort studies	Randomized trial or observational study with large effect	Individual randomized trial	Randomized trial
Local non-random sample	Non-consecutive studies, or studies without reference standards	Cohort study or control arm of randomized trial	Non-randomized controlled cohort	Cohort study with sufficient long term follow up	Cohort study
Case-series	Case-control studies or poor reference standard	Case-series or case- control studies, or poor quality cohort study	Case-series, case-control studies, or historically controlled studies	Case-series, case-control, or historically controlled studies	Case-series, case-control, or historically controlled studies
n/a	Mechanism-based reasoning	n/a	Mechanism-based reasoning	Mechanism-based reasoning	Mechanism-based reasoning

Figure 1. *The hierarchy of evidence.*

pling, lack of scientific insight, and poor understanding of topic content" [7]. These flaws and others are associated with many reviews. In general, only a reader familiar with the literature in question and the methodology for conducting high quality reviews will be able to identify unscientific reviews and their inappropriate conclusions.

Systematic reviews provide the reader with a critical source of information, and as clinicians whose time is a commodity, dentists may understandably turn to systematic reviews to guide practice. However, for the reasons mentioned earlier in this paper, a dentist's ability to evaluate the methodological quality of a systematic review should form the basis of a decision relating to the selection of a review to guide practice. Numerous tools have been developed since the first Quality of Reporting of Meta-analyses (QUOROM) conference was held in 1996. The major outcome of the conference was the development of QUOROM, the first reporting guideline. It was created to address the increasing quantity and varying quality of systematic reviews and meta-analyses [8]. It provided a check-list and a flow diagram for use in assessing

the quality of reporting on meta-analysis. In 2004, a review demonstrated that a total of 26 tools to evaluate systematic reviews had been developed since QUOROM. However, there are multiple shortcomings with the majority of them [9]. These shortcomings led to the development of an evaluation tool for the "Assessment of Multiple Systematic Reviews" (AMSTAR) (published in 2007) [10]. AMSTAR has only been tested for systematic reviews of interventions.

This paper describes the items within AMSTAR, a recently developed comprehensive evaluation tool that enables clinicians to assess effectively and efficiently results from systematic reviews as reliable, questionable or unreliable. It aims to highlight the aspects of systematic review methodology that influence its overall quality.

The AMSTAR tool

As mentioned previously, AMSTAR has been developed to evaluate the methodological quality of systematic reviews [10]. It comprises 11 concise criterion items (*Figure 2*); each item is given a score of 1 if the specific criterion is met, or a score

I. Was an 'apriori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

---

3. Was a comprehensive literature search performed?
At least two 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 searces 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.

---

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.

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

---

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.

---

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 in 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.

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 explicity stated in formulating recommendations.

---

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-squared 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?).

---

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

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

Figure 2. *AMSTAR-a measurement tool created to assess the methodological quality of systematic reviews. Shea BJ* et al. *(2007) [10] Open Access License BMC.*

of 0 if the criterion is not met, is unclear, or is not applicable. An overall score relating to review quality is then calculated (the sum of the individual item scores). AMSTAR characterises quality at three levels: 8 to 11 is high quality, 4 to 7 is medium quality, and 0 to 3 is low quality. Although scoring systems are controversial [11], the principles of the AMSTAR tool can be used to demonstrate aspects of systematic review methodology that influence the overall quality of a review. Each item of the AMSTAR tool will now be discussed in detail.

Item 1: Was a priori design provided?

A priori is Latin for "from the former" or "from before". In this context, it implies that the complete methodology to be employed for conducting a review has been predetermined. For Cochrane systematic reviews, the protocol is published as a standalone article in the Cochrane Library. The (International) Prospective Register of Systematic Reviews (PROSPERO) hosted by the Centre of Reviews and Dissemination is a relatively new international prospective register of systematic reviews, and provides a database of *a priori* systematic review designs, which are registered by the organisation (available at www.crd.york.ac.uk/prospero). The completed systematic review should be conducted according the *a priori* design and any divergence from the published protocol must be justified in the final review write up.

Item 2: Was there duplicate study selection and a data extraction?

Search results should be screened by at least two independent reviewers. This helps to prevent inappropriate inclusion or exclusion of articles and hence reduces bias in the selection of studies [12]. It has been suggested that the number of relevant articles found is increased by up to a third by using two reviewers instead of one [13]. In addition, data extraction from the included studies should be performed independently by the two reviewers. Any disagreements between them in relation to study selection or data extraction should be resolved by consensus. If the matter remains unresolved, a third party should be contacted to help reach a consensus. The procedure to be employed in such cases is generally reported in the protocol and detailed in the final review.

Item 3: Was a comprehensive literature search performed?

The search strategy used should be detailed in the protocol and the subsequent review. This should include details of the search terms used, and databases searched (including the years, for example MEDLINE 1966-April 2011). A minimum of two databases should be searched to ensure retrieval of studies irrespective of language and country of publication [12]. Language bias has been shown to influence publication patterns with positive outcomes more likely to be accepted into international English journals and negative outcomes in local journals [14]. Databases have geographical variations of their coverage, for example the Elsevier Medical Database (EMBASE) supplies good coverage to Western Europe (51%), whereas MEDLINE has a stronger position within North America (44%) [15]. The requirement for multiple databases has been highlighted and is encouraged to ensure that skewing of results is prevented through inadvertent exemption of valid studies [16].

All searches should be supplemented by consulting current content experts, reviews, textbooks, specialised registers, and by reviewing the references in the studies retrieved. An attempt at searching the "grey literature" and conference proceedings should also be made. In addition, if relevant journals are not indexed in the relevant databases they should be hand-searched.

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

Reports should be sought regardless of their publication type. Examples of grey literature include: conference abstracts, research reports, book chapters, unpublished data, dissertations, policy documents and personal correspondence [17]. Papers may not be published for a number of reasons; they may be written to support grant applications, inform funding parties of research results, address scientific concerns quickly, and the material may be distributed before or without being formally published [18]. The importance of grey literature is highlighted in a study of the antidepressant reboxetine, in which a pharmaceutical company withheld unpublished data, causing inconclusive outcomes over its safety, which was later found after the publication of the grey literature [19].

The authors should also state whether or not they excluded any reports, based on their publication status, language etc.

Item 5: Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided; the reasons for excluding any studies should also be provided. This shows transparency about the decision process employed by the authors, and it allows readers to decide for themselves whether they agree with the author's judgement on exclusion/inclusion or not. Without the exclusions being specified, publication bias is introduced because the reason for their exemption in unknown.

Item 6: Were the characteristics of the included studies provided?
Data from studies included in a review should be provided, ideally in an aggregated form such as a table. The data should comprise: author details, the country the study was conducted in, the year of publication, the number of participants involved, the interventions, any comparisons and the outcomes. The range of characteristics in all the studies analysed (e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases) should be reported. The presentation of characteristics in a table format facilitates direct comparison of the included studies and therefore it is convenient and reader friendly. This provides transparency, which helps the reader judge the relevance and generalisability of results to their own patients.

Item 7: Was the scientific quality of the included studies assessed and documented?
The quality of a study can be reflected in the extent to which that study reduces or eliminates bias and ensures reproducibility in its methodology. Bias is a systematic error that may lead in varying magnitudes to an under- or overestimation of effect [5]. There are numerous tools available for assessing the methodological quality of studies [20]; one tool the authors are familiar with is the domain-based evaluation used in Cochrane reviews. Authors of a systematic review appraise six domains (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting and other sources of bias) as being "low", "high" or "unclear". The assessment for each study is then presented in a "risk of bias" table in the review and can also be show in a graphic form (see *Figures 3a* and *3b* for examples).

Item 8: Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the quality assessment and risk of bias should be considered in the analysis and the conclusions of the review, and be explicitly stated in formulating recommendations. This helps to prevent changing practice based on poor-quality studies and conversely helps support practice change when good-quality studies provide such evidence (given that results are generalisable).

Item 9: Were the methods used to combine the findings of studies appropriate?
In some studies, the results from several studies may be pooled (a meta-analysis). In doing this, there is a more powerful indication of the effect and there is an increase in the precision of the results due to a larger data set [5]. For the pooling of results to be accurate and for the correct method to be used, studies need to be combinable and so their homogeneity needs to be assessed (for example using a statistical test such as the chi-square test for homogeneity or the I^2 test for heterogeneity). Although studies may be statistically homogenous, clinical diversities may mean they are un-combinable. For example, studies may vary in their participant characteristics (e.g., patient age), interventions (e.g., drug doses/ routes of administration) and outcomes (e.g., method or time of outcome measurement). These factors can lead to inaccurate conclusions being drawn if they are not accounted for by methods such as subgroup analysis.

Item 10: Was the likelihood of publication bias assessed?
Publication bias is the tendency for articles to be published due to their strength of findings [22]. It also refers to any influence that results in a reduction of quality literature being published [23]. It is widely recognised that when compared to studies with negative findings, those studies with results that are statistically significant and indicate a successful intervention are more likely to be published in high impact factor journals, and cited by others [24]. Statistical tests such as the funnel plot (which identifies a link between study size and effect of the intervention) and Egger regression are used to analyse a variety of factors that can cause publication bias. The outcome of these tests is a score that relates to the probability of publication bias occurring. An assessment of publication bias should

Risk of bias	Interpretation	Within a study	Across studies
Low risk of bias.	Plausible bias unlikely to seriously alter the results.	Low risk of bias for all key domains.	Most information is from studies at low risk of bias.
Unclear risk of bias.	Plausible bias that raises some doubt about the results.	Unclear risk of bias for one or more key domains.	Most information is from studies at low or unclear risk of bias.
High risk of bias.	Plausible bias that seriously weakens confidence in the results.	High risk of bias for one or more key domains.	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.

Figure 3a. *Example of a "risk of bias" table for a single study (fictional). Bessell* et al. *(2011) [21].*

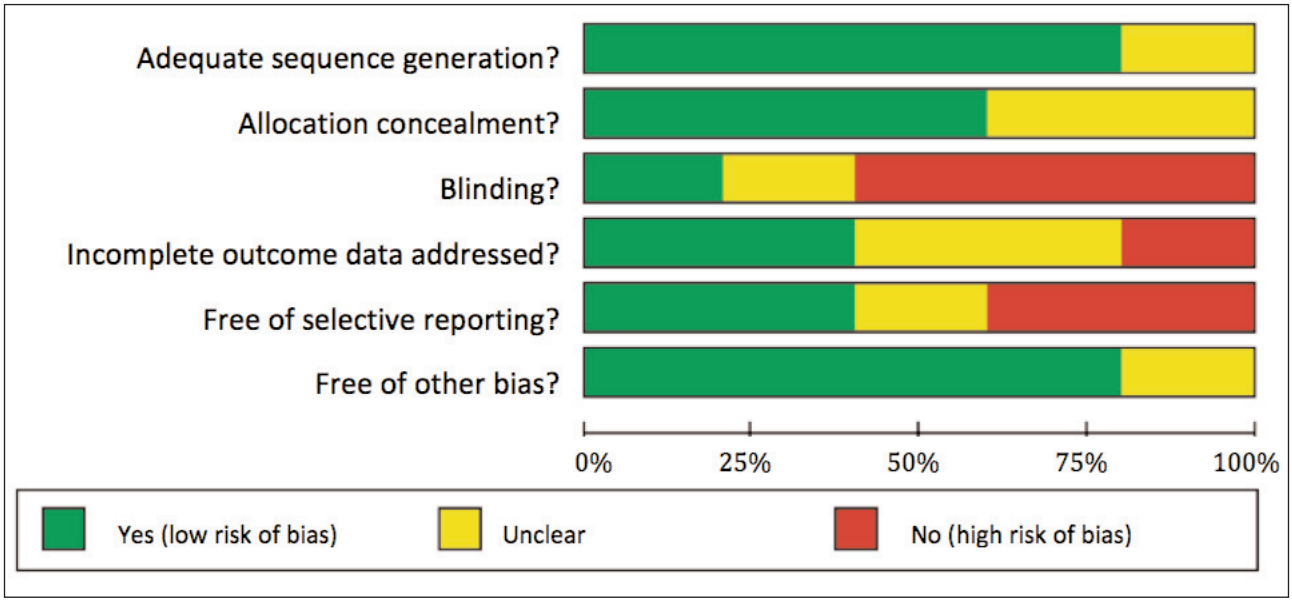


Figure 3b. *Example of a "risk of bias" graph taken from Bessell* et al. *(2011) [21].*

include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test). However, caution must be exercised when using these tests/aids as they are not without their own problems. For example, a funnel plot can appear asymmetric if the measure of effect is incorrect, or if there are differences due to effect size between large and small studies. This can lead to the incorrect conclusion that publication bias is prevalent. When carrying out any of these statistical tests, it is also important that there are sufficient numbers of studies to support the data produced.

Item 11: Was any conflict of interest stated?

Sources of support should be clearly acknowledged for both the systematic review and the included studies. There is evidence to support an association between some industry-supported systematic reviews and favourable outcomes of their products, resulting in product bias [25-28].

Discussion

Dentistry has historically been an empirical science in which experiment and expertise took precedence over research in influencing clinical decision making [29]. However, in modern day dentistry there is an ever-increasing and rapidly growing body of evidence. Clinicians can find themselves overwhelmed with the advent of new materials and techniques. These factors, combined with pressures from patients, manufacturers and the increasing importance of medico-legal issues, mean that evidence-based practice is now becoming the mainstream source of the decision-making process [19]. Evaluating this evidence is therefore imperative to daily practice and the development of clinical dentistry.

In light of this, the items of AMSTAR in evaluating the quality of systematic reviews have been described. AMSTAR provides clearly defined criteria that are quick and easy to follow. One study demonstrated that although 29% of general dental practitioners (GDPs) could not understand and use terms associated with evidence-based practice, 87% of them reported that they changed their practice after reading articles [30]. The influence of evidence-based practice on practice is apparent. However, not all systematic reviews are relevant to practice or have design methods that would lead to clinical change, AMSTAR and similar tools help to reveal methodologically sound systematic reviews.

AMSTAR has been proven to be a reliable through kappa analysis (inter-rater reliability was high at k=0.70) and a valid tool when compared with two other validated systematic review evaluation tools (QQAQ and Sacks' instrument) [31]. AMSTAR is an efficient tool, as on average the time taken to use it is 10-15 minutes, which is manageable in a time-pressured setting [20]. It provides a summary score, which is helpful for clinicians making decisions. Nevertheless, this can lead to masking of the specific strengths and weaknesses of an individual systematic review.

Reporting guidelines have been updated in the form of Preferred Reporting Items of Systematic reviews and Meta-Analyses (PRISMA) [32], and other methodological evaluation tools such as the *Critical Appraisal Skills Programme (CASP) of Systematic Reviews* [33], and Oxman and Guyatt (1991) [34]. However, in the opinion of the authors of this paper, the use of these tools is more time consuming.

All in all, AMSTAR provides a basic and effective method of evaluating systematic reviews for busy clinicians to ascertain the methodological quality of systematic reviews. The use of simple tools like AMSTAR helps to remove barriers such as the need for in-depth knowledge of research methodology in the clinician's pursuit of evidence-based dentistry within the context of everyday practice.

Conclusion

Assessing the quality of a systematic review has a crucial role in implementing evidence-based dentistry. The items in the AMSTAR tool that demonstrate the aspects of systematic review methodology that are influential to a review's overall quality have been described. Even if the AMSTAR tool is not adopted for use by readers of this journal, the authors hope that they have increased the understanding of quality assessment for systematic reviews.

The authors would like to point interested readers towards the *Cochrane Handbook of Systematic Reviews of Interventions* [3]: an invaluable resource for use during the design and conducting systematic reviews.

Funding

The authors confirm that no sources of funding were obtained. Mohammad Owaise Sharif is supported by a National Institute for Health Research (NIHR) Academic Clinical Fellowship.

Contributions of each author

- MOS conceived the idea for the paper, designed, supervised and completed the final manuscript.
- FA drafted the final manuscript.
- FNJS drafted the final manuscript.
- HA drafted the initial stages of the manuscript.

Statement of conflict of interest

The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, The University of Manchester or the Department of Health, UK.

References

1. The Cochrane Collaboration. Cochrane Reviews [website]. Accessed (2011 May 18) at: http://www.cochrane.org/cochrane-reviews

2. Oxford Centre for Evidence-Based Medicine (OCEBM) [website]. OCEBM Levels of Evidence Working Group. The Oxford 2011 Levels of Evidence. http://www.cebm.net/index.aspx?o=5653

3. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0. Chichester: Wiley; 2011. Accessed (2011 May 18) at: www.cochrane-handbook.org

4. Cochrane Prognosis Methods Group [website]. Accessed (2012 Mar 4) at: http://prognosismethods.cochrane.org/

5. The Cochrane Collaboration. Diagnostic Test Accuracy Working Group [website]. Accessed (2012 Mar 4) at: http://srdta.cochrane.org/

6. Jin W, Yu R, Li W, Youping L, Ya L, Min Z, *et al*. The reporting quality of meta-analyses improves: A random sampling study. *Journal of Clinical Epidemiology.* 2008; **61**: 770-775.

7. Spångberg LS. Systematic reviews in endodontics—examples of GIGO? *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*. 2007; **103**: 723-724.

8. 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. Quality of Reporting of Meta-analyses. *Lancet*. 1999; **354**: 1896-1900.

9. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology.* 2004; **4**: 22.

10. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, *et al*. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology.* 2007; **7**: 10.

11. Féry-Lemonnier E, Landais P, Loirat P, Kleinknecht D, Brivet F. Evaluation of severity scoring systems in ICUs—translation, conversion and definition ambiguities as a source of inter-observer variability in Apache II, SAPS and OSF. *Intensive Care Medicine*. 1995; **21**: 356-360.

12. Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine. 2002*; **21**: 1635-1640.

13. Cooper H, Ribble RG. Influences on the outcome of literature searches for integrative research reviews. *Knowledge.* 1989; **10**: 179-201.

14. Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet*. 1997; **350**: 326-329.

15. EMBASE [website]. FAQ. How is the coverage of Embase broken down geographically? Accessed (2012 Mar 12) at: http://www.embase.com/info/faq/how-coverage-embase-broken-down-geographically

16. Betrán AP, Say L, Gülmezoglu AM, Allen T, Hampson L. Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. *BMC Medical Research Methodology.* 2005; **5**: 6.

17. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*. 2007; Apr 18; (2): MR000010.

18. European Association for Grey Literature Exploitation (EAGLE). *SIGLE Manual Part 2: Subject Category List.* 3rd ed. The Hague: EAGLE; 1991.

19. Wieseler B, McGauran N, Kaiser T. Finding studies on reboxetine: a tale of hide and seek. *British Medical Journal.* 2010; **341**: c4942.

20. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*. 1995; **16**: 62-73.

21. Bessell A, Hooper L, Shaw WC, Reilly S, Reid J, Glenny AM. Feeding interventions for growth and development in infants with cleft lip, cleft palate or cleft lip and palate. *Cochrane Database of Systematic Reviews*. 2011; (2): CD003315. doi: 10.1002/14651858.CD003315.pub3

22. Dickersin K. The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association.* 1990; **263**: 1385-1389.

23. Chalmers T, Frank C, Reitman D. Minimising the three stages of publication bias. *Journal of the American Medical Association.*1990; **263**: 1392-1395.

24. Sterne JAC, Egger M, Moher D, editors. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0. Chichester: Wiley; 2011

25. Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Medicine*. 2007; **4**: 184.

26. Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *British Medical Journal*. 2006; **333**: 782.

27. Jørgensen AW, Maric KL, Tendal B, Faurschou A, Gøtzsche PC. Industry-supported meta-analyses compared with meta-analyses with non-profit or no support: differences in methodological quality and conclusions. *BMC Medical Research Methodology.* 2008; **9**: 60.

28. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *British Medical Journal*. 2007; **335**: 1202-1205.

29. Bader JD, Shugars DA. Variation, treatment outcomes, and practice guidelines in dental practice. *Journal of Dental Education*. 1995; **59**: 61-95.

30. Iqbal A, Glenny AM. General dental practitioners' knowledge of and attitudes towards evidence based practice. *British Dental Journal*. 2002; **193**: 58.

31. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, *et al* AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology.* 2009; **62**: 1013-1020.

32. PRIMSA [website]. Transparent Reporting of Systematic Reviews and Meta-Analyses. Accessed (2011 Jun 5) at: http://www.prisma-statement.org/endorsing_prisma.htm

33. NHS Solutions for Public Health. Critical Appraisal Skills Programme [website]. Accessed (2011 Jun 5) at: http://www.sph.nhs.uk/what-we-do/public-health-workforce/resources/critical-appraisals-skills-programme

34. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology.* 1991; **44**: 1271-1278.