Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Systematic reviews with language restrictions and no author contact have lower overall credibility: a methodology study

This article was published in the following Dove Press journal:
Clinical Epidemiology
31 March 2015
Number of times this article has been viewed

Zhen Wang[1–3]
Juan P Brito[4]
Apostolos Tsapas[5]
Marcio L Griebeler[4]
Fares Alahdab[1,3]
Mohammad Hassan Murad[1,3,6]

[1]Robert D and Patricia E Kern Center for the Science of Health Care Delivery, [2]Division of Health Care Policy and Research, Department of Health Sciences Research, [3]Knowledge and Evaluation Research Unit, [4]Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, MN, USA; [5]Aristotle University of Thessaloniki, Thessaloniki, Greece; [6]Division of Preventive, Occupational and Aerospace Medicine, Mayo Clinic, Rochester, MN, USA

Background: High-quality systematic reviews (SRs) require rigorous approaches to identify, appraise, select, and synthesize research evidence relevant to a specific question. In this study, we evaluated the association between two steps in the conduct of an SR – restricting the search to English, and author contact for missing data – and the overall credibility of a SR.

Methods: All SRs cited by the Endocrine Society's Clinical Practice Guidelines published from October 2006 through January 2012 were included. The main outcome was the overall A Measurement Tool to Assess Systematic Reviews (AMSTAR) score, as a surrogate of SR credibility. Nonparametric Kruskal–Wallis tests and multivariable linear regression models were used to investigate the association between language restriction, author contact for missing data, and the overall AMSTAR score.

Results: In all, 69 SRs were included in the analysis. Only 31 SRs (45%) reported searching non-English literature, with an average AMSTAR score of 7.90 (standard deviation [SD] =1.64). SRs that reported language restriction received significantly lower AMSTAR scores (mean =5.25, SD =2.32) ($P<0.001$). Only 30 SRs (43%) reported contacting authors for missing data, and these received, on average, 2.59 more AMSTAR points (SD =1.95) than those who did not ($P<0.001$). In multivariable analyses, AMSTAR score was significantly correlated with language restriction (beta =–1.31, 95% confidence interval [CI]: –2.62, –0.01, $P=0.05$) and author contact for missing data (beta =2.16, 95% CI: 0.91, 3.41, $P=0.001$). However, after adjusting for compliance with reporting guidelines, language restriction was no longer significantly associated with the AMSTAR score.

Conclusion: Fewer than half of the SRs conducted to support the clinical practice guidelines we examined reported contacting study authors or searched non–English literature. SRs that did not conduct these two steps had lower quality scores, suggesting the importance of these two steps for overall SR credibility.

Keywords: evidence-based medicine, research design, validity, quality of evidence

Introduction

Systematic reviews (SRs) are the foundation of evidence-based medicine and the best way to summarize the highest level of evidence that guides clinicians, patients, and other stakeholders in decision making. With the intention to minimize bias in the selection and appraisal of individual studies, SRs employ rigorous approaches to identify, appraise, select, and synthesize research evidence relevant to a specific question.

However, like other types of study design, not all SRs are credible. An SR with lower credibility can distort evidence. Several authorities and organizations have provided guidance to improve the quality of conducting SRs,[1–3] such as the Cochrane Collaboration, the Agency for Healthcare Research and Quality (AHRQ), and the

Correspondence: Zhen Wang
Robert D and Patricia E Kern Center for the Science of Health Care Delivery, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA
Tel +1 507 538 6153
Fax +1 507 538 0850
Email wang.zhen@mayo.edu

Institute of Medicine (IOM). A user's guide for interpreting and applying the results of SRs has also been developed.[4] Checklists and instruments specifically designed to appraise SRs have also been developed.[5–11] For example, A Measurement Tool to Assess Systematic Reviews (AMSTAR) is one of the commonly used tools. It was developed in 2007 by combining items from existing tools, using experts' input and exploratory factor analysis to finish with an 11-item instrument.[5] The tool was found to have reasonable reliability and validity.[12,13] However, the tool does not include two features recommended to improve the credibility of SRs, specifically, the contacting of authors of included studies in SRs for additional data/verification of extracted data, and inclusion of all languages in the literature search.[2,14]

In this study, we aimed to evaluate the association between two steps in the conduct of an SR – language restrictions and author contact for missing data – and the overall credibility of a SR as measured by AMSTAR. We hypothesized that inclusion of these two procedures, which are not part of AMSTAR, might nevertheless be associated with higher AMSTAR scores and increased overall credibility, providing additional rationale to conduct these two steps.

Methods

Data sources

The details of the data sources were described in a previous report.[15] Briefly, we identified all SRs cited by the Endocrine Society's Clinical Practice Guidelines from October 2006 through January 2012. Diagnostic SRs, SRs of preclinical studies, and SRs without meta-analysis were excluded.

We extracted data describing the characteristics of each SR including the eleven items of AMSTAR (statement of priori design, duplicate study selection and data extraction, comprehensive literature search, status of publication used as an inclusion criterion, list of studies, characteristics of the included studies, scientific quality of the included studies assessed and documented, scientific quality of the included studies used appropriately, appropriate methods used to combine the findings of studies, publication bias assessed, and statements of conflict of interest).[5] We also extracted data on the two SR steps we hypothesized to be associated with AMSTAR score (language restriction and author contact for missing data).Two independent reviewers extracted study details from the full text of the included SRs. All conflicts between the two reviewers were resolved through discussions and consensus. We reached near perfect agreement between the two reviewers as measured by chance-adjusted interrater agreement (Cohen's kappa =0.91).[16] We also hypothesized that certain confounders might affect this association, such as SR funding source, study design of included studies, whether the SR followed a specific SR reporting guideline or statement, the impact factor of the journal,[17] and the number of published manuscript pages. Some empiric evidence supports the association of these variables with AMSTAR score, perhaps reflecting that better reporting of SRs leads to higher AMSTAR scores. For example, an evaluation of SRs in gastroenterology suggested a significant association between the number of published manuscript pages of an SR and SR quality (the longer the manuscript, the better the SR).[8] Also, SRs in endocrinology that summarized randomized controlled trials (RCTs) were found to have higher quality than those summarizing observational studies.[15] Both observations are potentially associated with the quality of reporting in SRs rather than the validity/credibility of the SR findings. One example is that the longer manuscript allows more details in SR manuscript, which can lead to a higher AMSTAR score.

Statistical analysis

The main outcome of interest was the AMSTAR score, as a surrogate of SR credibility. SRs received 1 point for a "yes" answer for each AMSTAR item. The overall AMSTAR score was calculated by aggregating the total points a SR received, with a maximum of 11 points. Language restriction was determined by whether the SR restricted the literature search to studies published in English. This was categorized as: yes, no, or unknown. Author contact for missing data was categorized as: yes, no, or unknown. We used the following categories for the confounders: funding source of SRs (nonprofit, for profit, or unknown), design of the included studies (RCTs included, no RCTs, or unknown), whether the SR followed a reporting guideline (yes or no), impact factor of the journal in which the SR was published ($\leq$6 or $>$6), and number of the published manuscript pages ($\leq$10 pages or $>$10 pages). We conducted descriptive analyses to evaluate the associations of each variable and the overall AMSTAR score. A nonparametric Kruskal–Wallis test was used to test the significant difference for categorical variables. We used multivariable analyses to evaluate the strength of the association between language restriction, author contact, and the AMSTAR score. We excluded the category "unknown" from the analysis. Due to the small number of SRs included in this study, we first constructed a multivariable linear regression model by including only language restriction and author contact. Then, we added each of the confounders, including funding source, study design of included studies, whether the

SR followed a reporting guideline, the impact factor of the published journal, and the number of published manuscript pages, one at a time to evaluate the robustness of the findings. All statistical analyses were conducted using STATA version 12.1 (StataCorp, College Station, TX, USA).

Results

A total of 69 SRs met the inclusion criteria and were included in the analysis. The included SRs were published between 1988 and 2012. The clinical areas were related to the pituitary–gonad–adrenal axis (42%), metabolism (30%), diabetes (16%), bone metabolism (7%), and other endocrinology topics (4%).

The mean AMSTAR score of the SRs was 6.36, with a standard deviation (SD) of 2.48. Table 1 shows the AMSTAR score by each tested variable. A total of 31 SRs (45%) reported having no language restrictions, 16 (23%) reported language restrictions, and 22 SRs (32%) did not clarify whether language restriction was used. The SRs without language restrictions reported the highest AMSTAR score (mean =7.90, SD =1.64), and the difference was significant ($P<0.001$). A total of 30 SRs (43%) reported contacting authors for missing data and received an average AMSTAR score of 7.73 (SD =1.95), while 35 SRs (51%) did not contact author and received an average score of 5.14 (SD =2.21). The AMSTAR score difference was 2.59 points ($P<0.001$). With regards to the confounders, we found significantly higher AMSTAR scores in SRs funded by nonprofit sources than for those funded by profit sources (7.20 vs 6.60) ($P<0.001$). A higher AMSTAR score was also found in SRs that included RCTs than in those without RCTs (7.11 vs 4.90) ($P=0.002$), in SRs that used SR-specific reporting guideline (7.89 vs 4.79) ($P<0.001$), in those published in higher impact journals (7.02 vs 5.13) ($P=0.02$), and in those with more manuscript pages (7.14 vs 5.80) ($P=0.03$).

In the multivariable analyses (Table 2), the AMSTAR score was significantly correlated with language restriction (beta =–1.31, 95% confidence interval [CI]: –2.62, –0.01, $P=0.05$) and author contact for missing data (beta =2.16, 95% CI: 0.91, 3.41, $P=0.001$). By adding the confounders one at a time in the analyses, we found the significant associations remained for all except one. After adjusting for compliance with reporting guidelines, language restriction was no longer significantly associated with the AMSTAR score (beta =–0.89, 95% CI: –2.13, 0.36, $P=0.16$). However, the association with author contact continued to be significant (beta =1.67, 95% CI: 0.46, 2.88, $P=0.01$).

Table 1 Univariate analysis of language restriction, author contact, confounders, and the overall AMSTAR score

Variables	Categories	Studies (N=69), n (%)	AMSTAR score, mean (SD)	*P*-value
Language restriction	Yes	16 (23%)	5.25 (2.32)	<0.001
	No	31 (45%)	7.90 (1.64)	
	Unknown	22 (32%)	5.00 (2.41)	
Author contact for missing data	Yes	30 (43%)	7.73 (1.95)	<0.001
	No	35 (51%)	5.14 (2.21)	
	Unknown	4 (6%)	6.75 (3.30)	
Funding source of SRs	Nonprofit	44 (64%)	7.20 (2.03)	<0.001
	For profit	5 (7%)	6.60 (2.51)	
	Unknown	20 (29%)	4.45 (2.42)	
Design of included studies	RCTs included	46 (67%)	7.11 (2.23)	0.002
	No RCTs	21 (30%)	4.90 (2.36)	
	Unknown	2 (3%)	4.52 (2.12)	
Reporting guideline	Yes	35 (51%)	7.89 (0.29)	<0.001
	No	34 (49%)	4.79 (0.37)	
Impact factor of the published journal	≤6	25 (36%)	5.13 (2.76)	0.02
	>6	44 (64%)	7.02 (2.09)	
Number of pages in the published manuscript	≤10 pages	40 (58%)	5.80 (2.49)	0.03
	>10 pages	29 (42%)	7.14 (2.29)	

Abbreviations: AMSTAR, A Measurement Tool to Assess Systematic Reviews; RCTs, randomized controlled trials; SD, standard deviation; SRs, systematic reviews.

Discussion

Main findings

In this study, we evaluated the association between two recommended steps for the conduct of SRs – language restriction and author contact for missing data – and the overall credibility of an SR, by evaluating 69 SRs cited by the Endocrine Society's Clinical Practice Guidelines from October 2006 through January 2012. We found significant associations between language restriction, author contact for missing data, and the overall AMSTAR score, a surrogate for credibility and rigor of SRs. Another important finding was that less than a half of the SRs contacted authors for missing data or had an unrestricted language search.

Table 2 Multiple linear regression analysis of quality indicators and AMSTAR score

Quality indicators	Coefficient	95% confidence interval	*P*-value
Language restriction	–1.31	–2.62, –0.01	0.05
Author contact for missing data	2.16	0.91, 3.41	0.001

Abbreviation: AMSTAR, A Measurement Tool to Assess Systematic Reviews.

Strengths and limitations

In this study, we specifically evaluated SRs that were cited in clinical practice guidelines. These are the evidence summaries most proximal to implementation and important from a perspective of evidence users (patients and policy makers). We identified SRs and extracted data in duplicate, with excellent interreviewer agreement. We attempted to control for some confounders for which we had empirical evidence of an associations with the AMSTAR score. At last, to our knowledge, we are the first to compare the two SR steps with an overall quality measure, though others have compared restricted language search to nonrestricted search in terms of study identification and retrieval.[18–25]

This study suffers several important limitations. First, we used the AMSTAR score as the proxy of the credibility of SRs. Although AMSTAR was shown to have good reliability and validity, like other quality instruments, it is prone to mixing two concepts: process credibility (eg, how well the SR was conducted) and reporting quality; thus, it is not a "gold standard". Second, due to the small number of SRs included in the analyses, we were not able to adjust for all of the confounders at the same time. Other cofounders may exist that might affect our findings. Third, we used the overall AMSTAR score realizing that the weighting of the different items in AMSTAR are not equal in terms of importance and would likely differ per clinical question.

Implications for research

There is increasing recognition that high-quality SRs are critical in providing valid, reliable, and high-quality evidence in clinical decision making. Guidelines developed by organizations such as the Cochrane Collaboration, the AHRQ, and the IOM, provide rigorous and detailed information on how to conduct SRs. Although these organizations generally require unrestricted language search, only 45% of the SRs included in this study actually complied. The AHRQ manual recommends contacting the author for missing information, and the Cochrane collaboration suggests this and highlights some of its challenges.[1,2]

Quality assessment tools, such as AMSTAR, cannot contain all possible quality indicators. Instrument developers have to be parsimonious and practical in choosing items that will lead to a useful tool. Therefore, we are not suggesting adding language restriction and author contact to AMSTAR. Nor are we suggesting that our results validate AMSTAR. Rather, we demonstrated that these two important steps are associated with higher scores of a surrogate of SR credibility, and suggest that systematic reviewers should try to perform these steps whenever feasible.

Conclusion

Less than half of the SRs conducted to support the clinical practice guidelines we examined contacted study authors or searched non-English literature. SRs that did not conduct these two steps had lower quality scores, suggesting the importance of these two steps for overall SR credibility.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Chang SM. The Agency for Healthcare Research and Quality (AHRQ) effective health care (EHC) program methods guide for comparative effectiveness reviews: keeping up-to-date in a rapidly evolving field. *J Clin Epidemiol*. 2011;64(11):1166–1167.
2. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* Chichester: John Wiley & Sons Ltd; 2008.
3. Eden J, Levit L, Berg A, Morton S, editors; Committee on Standards for Systematic Reviews of Comparative Effectiveness Research; Institute of Medicine (US). *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: National Academies Press; 2011.
4. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *Jama*. 2014;312(2):171–179.
5. 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.
6. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*. 1991;44(11):1271–1278.
7. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120(8):667–676.
8. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: an update. *Mt Sinai J Med*. 1996;63(3–4):216–224.
9. Auperin A, Pignon JP, Poynard T. Review article: critical review of meta-analyses of randomized clinical trials in hepatogastroenterology. *Aliment Pharmacol Ther*. 1997;11(2):215–225.
10. Beck CT. Use of meta-analysis as a teaching strategy in nursing research courses. *J Nurs Educ*. 1997;36(2):87–90.
11. Smith AF. An analysis of review articles published in four anaesthesia journals. *Can J Anaesth*. 1997;44(4):405–409.
12. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009;62(10):1013–1020.
13. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One*. 2007;2(12):e1350.
14. Mullan RJ, Flynn DN, Carlberg B, et al. Systematic reviewers commonly contact study authors but do so with limited rigor. *J Clin Epidemiol*. 2009;62(2):138–142.
15. Brito JP, Tsapas A, Griebeler ML, et al. Systematic reviews supporting practice guideline recommendations lack protection against bias. *J Clin Epidemiol*. 2013;66(6):633–638.
16. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *Cmaj*. 2004;171(11):1369–1373.

17. Web of ScienceTM. [homepage on the Internet]. 2012 Journal Citation Reports. Thomson Reuters. Available from: http://www.isiwebofknowledge.com. Accessed March 17, 2015.
18. 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(9074):326–329.
19. Heres S, Wagenpfeil S, Hamann J, Kissling W, Leucht S. Language bias in neuroscience – is the Tower of Babel located in Germany? *Eur Psychiatry*. 2004;19(4):230–232.
20. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials*. 1998;19(2):159–166.
21. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess*. 2003;7(1):1–76.
22. Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol*. 1995;48(1):159–163.
23. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol*. 2002;31(1):115–123.
24. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol*. 2000;53(9):964–972.
25. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. *J Clin Epidemiol*. 2005;58(8):769–776.