REVIEW

Assessment for Risk of Bias in Systematic Reviews and Meta-Analyses in the Field of Hepatology

Gaeun Kim*, Youn Zoo Cho[†], and Soon Koo Baik[†]

Department of Nursing, Keimyung University College of Nursing, Daegu, and [†]Department of Internal Medicine, Wonju Severance Christian Hospital, Yonsei University Wonju College of Medicine, Wonju, Korea

A systematic review (SR) provides the best and most objective analysis of the existing evidence in a particular field. SRs and derived conclusions are essential for evidence-based strategies in medicine and evidence-based guidelines in clinical practice. The popularity of SRs has also increased markedly in the field of hepatology. However, although SRs are considered to provide a higher level of evidence with greater confidence than original articles, there have been no reports on the quality of SRs and meta-analyses (MAs) in the field of hepatology. Therefore, we performed a quality assessment of 225 SRs and MAs that were recently published in the field of hepatology (January 2011 to September 2014) using A MeaSurement Tool to Assess systematic Reviews (AMSTAR). Using AMSTAR, we revealed both a shortage of assessments of the scientific quality of individual studies and a publication bias in many SRs and MAs. This review addresses the concern that SRs and MAs need to be conducted in a stricter and more objective manner to minimize bias and random errors. Thus, SRs and MAs should be supported by a multi-disciplinary approach that includes clinical experts, methodologists, and statisticians. (Gut Liver 2015;9:701-706)

Key Words: Systematic review; Meta-analysis; Assess the Methodological quality of systematic review; Hepatology

INTRODUCTION

Currently, with the immense growth of electronic publications, the volume of medical literature that is published yearly exceeds the volume that can be reviewed by experts, and studies with conflicting results on the same topic are common. This situation can make it difficult to draw definitive conclusions; thus, systematic reviews (SRs) provide the best and most trustworthy objective analysis of the existing evidence.[1]

SRs require a process that involves the definition of the review question, the search for studies, the selection of studies, and the collection (i.e., retrieval) of data. This process is necessary to assess the risk of bias in the included studies, analyze the data, perform a meta-analysis (MA), and interpret the results. Each step should be conducted independently by at least two researchers. The validity of the selected studies should be assessed via an evaluation of the risk of bias in the study results (i.e., the risk that the authors will overestimate or underestimate the true intervention effect). Studies in which the conclusions are not based on valid and objective evidence or in which the validity of the methodology is not robust cannot provide reliable answers to the questions addressed in a SR. Various types of tools are used to evaluate the quality of a given study. Quality assessment tools include the risk of bias (e.g., Cochrane Library),[2] checklists Scottish Intercollegiate Guidelines Network (SIGN),[3] Quality Assessment of Diagnostic Accuracy Studies, and the Newcastle-Ottawa Scale and so on.

Recently, many SRs and MAs have been published. However, negative opinions are prevalent regarding the risk of bias and the quality of the research. It is important to ensure that studies are based on methodological principles. Therefore, some evidence-based practice development centers regularly perform these tasks. SRs and MAs can be conducted with the processes described above, and they can refer either to the "reporting guidelines" presented by the Preferred Reporting Items for Systematic Review and Meta-Analysis Group[4] or A MeaSurement Tool to Assess systematic Reviews (AMSTAR).[5] Because SRs and MAs are essential for evidence-based medicine in the decision-making process in the public policy realm, a strict and objective

research process and methodology are necessary. Thus, a systematic literature search and an assessment of the risk of bias in the selected literature are needed.[1]

In this review, we discuss a general descriptive assessment of the methodological quality of SRs and MAs published in the field of hepatology from 2011 to 2014 using AMSTAR.

HOW CAN THE RISK OF BIAS IN SRs AND MAs BE EVALUATED?

A search was performed for SRs and MAs published in the field of hepatology using the ISI Web of Knowledge site, Ovid-MEDLINE, PubMed, and Google, among others. The inclusion criteria included the following: (1) articles in 74 SCI or SCIE journals with the subject categories of gastroenterology and hepatology; (2) hepatology-related SRs or MAs; (3) studies published between January 2011 and September 2014; and (4) studies with full text available. The exclusion criteria were as follows: (1) SRs or MAs of laboratory studies or animal experiments and (2) studies not published in English.

The methodological quality of the included studies was assessed by AMSTAR. The most commonly used assessment tools for the methodological quality of SRs and MAs were AMSTAR and the SIGN checklist. AMSTAR was developed in 2007, and 11 items were selected after consideration of the Overview Quality Assessment Questionnaire (10 items) and Sack's checklist (24 items).[6] The SIGN checklist was amended in 2013 and is based on AMSTAR. Because the two tools feature almost identical questions, we chose AMSTAR. This tool is an 11-item questionnaire that can be used to assess the methodological quality of SRs by assessing the presence of the following: an a priori design, duplicate study selection and data extraction, a comprehensive literature search, the use of publication status as an inclusion criterion, a list of included/excluded studies, the characteristics of included studies, a documented assessment of the scientific quality of included studies, appropriate consideration of the scientific quality in the formation of conclusions, the appropriate use of methods to combine findings from multiple studies, an assessment of the likelihood of publication bias, and the documentation of conflict of interest. However, two of the 11 items were slightly modified as follows:

"Was an 'a priori' design provided?" was amended to "Was the research question (i.e., research purpose) clarified?" The reason for this modification is that the exclusion of an a priori design from a protocol is not unusual except in the case of publications in the Cochrane Library.

"Was the status of publication (i.e., grey literature) used as an inclusion criterion?" was modified to "Were inclusion/exclusion criteria reported clearly?" This item should be retrieved, but it is difficult to extract data from grey literature or from unpublished literature.

In the first step of the present study, one of the authors screened and retrieved eligible articles using a sensitive search strategy with broad inclusion criteria that were established a priori in the study. In the second step, both researchers independently reviewed the studies based on the full-text articles. The methodological quality of the SRs and MAs was independently assessed by both researchers using AMSTAR, a validated 11-item tool. Disagreements were resolved by discussion.

According to the AMSTAR criteria, we assigned a "yes" when a criterion was satisfied and a "no" when a criterion was not satisfied (Supplementary Data 1).

We researched the trends by publication year, disease category, and journal. The frequency of response to the 11 AMSTAR items and the overall quality of the studies were assessed. After we analyzed the results of the 11 AMSTAR items, we summarized the overall results to facilitate a comparison of all items. The results summary was based on the assignment of a "yes" response to the items. Scores were assigned as follows: ++++, percentage of "yes" items was 80% to 100% (nine questions); +++, percentage of "yes" items was 60% to 80% (seven to eight questions); ++, percentage of "yes" items was 40% to 60% (five to six questions); and +, percentage of "yes" items were less than 40% (four items or less). In one case, the answer to a question was "not applicable"; in that case, a total of 10 items were available.

CHARACTERISTICS OF SRs AND MAs FROM 2011 TO 2014

This study included 225 studies that reported results in the form of complete papers that were published during the period of 2011 to 2014. Among the 74 journals with subject categories of gastroenterology and hepatology, most belong to the SCI or SCIE indices.

With respect to the total number of SRs and MAs, 51, 50, 67, and 57 were published in 2011, 2012, 2013, and 2014 (September to present), respectively (Supplementary Data 2). The publication of these types of papers has continued to increase (Table 1). Thirty-two of the identified studies were published in *World J Gastroenterol*, which was the greatest number of studies published in a single journal. We also included studies published in the following journals: *Aliment Pharmcol Ther* (24 studies), *Hepatol Res* (17 studies), *HPB* (*Oxford*) (13 studies), *Hepatogastroenterology* (12 studies), *Eur J Gastroenterol Hepatol* (11 studies), *J Hepatol* (10 studies), and *Liver Int* (eight studies).

Considering only SRs and MAs published within the last 4 years, only one journal contained more than 30 studies, and most journals contained fewer than 10 studies. Regarding study type, a total of 59 SRs, 79 MAs, and 87 combined SRs and MAs were analyzed.

Table 1. Current Status by Year and Journals of Published Literature

Journal title	No. of studies				Total
	2011	2012	2013	Sep 2014	
1. *Hepatology*	1	2	1	–	4
2. *J Hepatol*	2	7	1	–	10
3. *Aliment Pharmacol Ther*	7	3	2	12	24
4. *Liver Int*	3	–	2	3	8
5. *Liver Transpl*	1	4	1	1	7
6. *J Gastroenterol Hepatol*	3	1	1	1	6
7. *J Viral Hepat*	3	1	2	2	8
8. *Dig Dis Sci*	3	2	2	2	9
9. *World J Gastroenterol*	8	9	7	8	32
10. *J Gastrointest Surg*	1	1	2	–	4
11. *J Hepatobiliary Pancreat Sci*	1	–	–	1	2
12. *Hepatol Res*	3	4	4	6	17
13. *Ann Hepatol*	1	–	2	1	4
14. *Eur J Gastroenterol Hepatol*	1	1	7	2	11
15. *BMC Gastroenterol*	1	2	4	–	7
16. *HPB (Oxford)*	1	4	3	5	13
17. *Colorectal Dis*	1	–	–	–	1
18. *J Dig Dis*	1	–	–	–	1
19. *J Gastrointestin Liver Dis*	1	–	–	–	1
20. *Hepat Mon*	4	–	3	1	8
21. *Hepatobiliary Pancreat Dis Int*	2	1	1	2	6
22. *Hepatogastroenterology*	2	5	5	–	12
23. *Best Pract Res Clin Gastroenterol*	–	1	–	–	1
24. *Can J Gastroenterol*	–	2	2	–	4
25. *Am J Gastroenterol*	–	–	1	–	1
26. *Gastrointest Endosc*	–	–	1	–	1
27. *J Clin Gastroenterol*	–	–	1	2	3
28. *Dig Liver Dis*	–	–	1	2	3
29. *Pancreatology*	–	–	1	–	1
30. *Hepatol Int*	–	–	1	1	2
31. *Scand J Gastroenterol*	–	–	1	–	1
32. *Clin Res Hepatol Gastroenterol*	–	–	2	2	4
33. *Gastroenterol Res Pract*	–	–	3	1	4
34. *Gastroenterol Nurs*	–	–	1	–	1
35. *Turk J Gastroenterol*	–	–	1	1	2
36. *Gastroenterology*	–	–	1	–	1
37. *Saudi J Gastroenterol*	–	–	–	1	1
Total	51	50	67	57	225

Journal titles are abbreviated in accordance with the style of Index Medicus.

CATEGORY OF LIVER DISEASES

When the papers were analyzed according to disease type, the most common disease was hepatocellular carcinoma (36.9%), followed by hepatitis C virus (HCV) (13.3%), cirrhosis and its subsequent complications (12%), hepatitis B virus (HBV) (9.3%), and nonalcoholic fatty liver disease (8.9%) (Fig. 1).
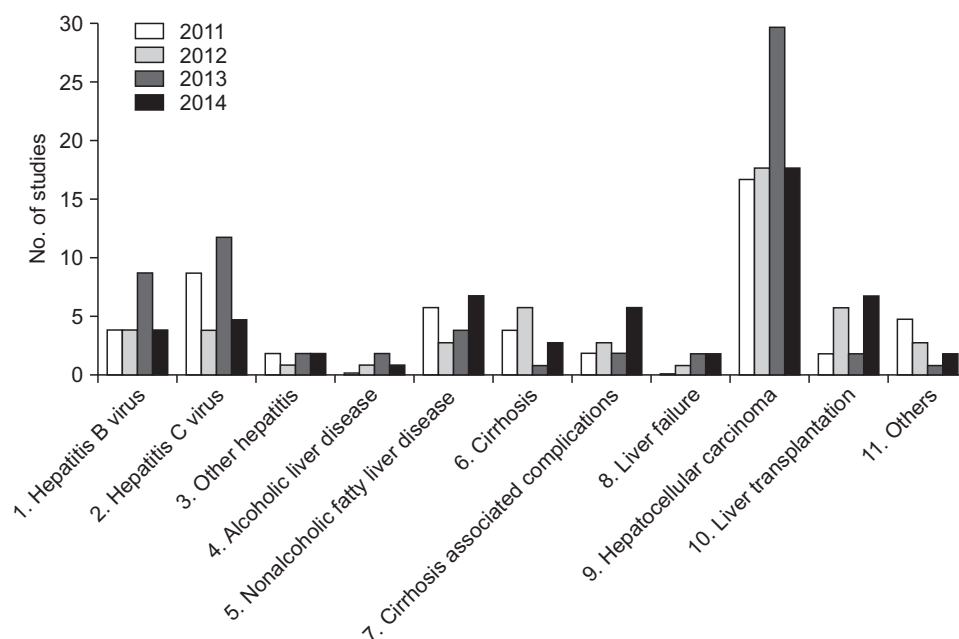
Fig. 1. Category of liver diseases.

Table 2. The Ratio of "Yes" Using the AMSTAR Checklist

Item	No. of studies (%), (n=225)			
	Yes	No	Cant's answer	Not applicable
Was an "a priori" design provided?	225 (100.0)	0	0	0
Was there duplicate study selection and data extraction?	183 (81.4)	41 (18.2)	1 (0.4)	0
Was a comprehensive literature search performed?	204 (90.7)	21 (9.3)	0	0
Was the status of publication (i.e., grey literature) used as an inclusion criterion?	205 (91.1)	20 (8.9)	0	0
Was a list of studies (included and excluded) provided?	213 (94.7)	12 (5.3)	0	0
Where the characteristics of the included studies provided?	203 (90.7)	21 (9.3)	0	0
Was the scientific quality of the included studies assessed and documented?	130 (57.8)	94 (41.8)	1 (0.4)	0
Was the scientific quality of the included studies used appropriately in formulating conclusions?	127 (56.4)	94 (41.8)	2 (0.9)	2 (0.9)
Were the methods used to combine the findings of studies appropriate?	155 (68.9)	10 (4.4)	0	60 (26.7)
Was the likelihood of publication bias assessed?	122 (54.2)	103 (45.8)	0	0
Was the conflict of interest stated?	182 (80.9)	43 (19.1)	0	0

AMSTAR, A MeaSurement Tool to Assess systematic Reviews.

ANALYSIS OF THE METHODOLOGICAL QUALITY

The methodological quality of each item was determined using the AMSTAR. The results for each item are as follows (Table 2):

(1) Was the research question (research purpose) clarified?

The SR or MA must have an established protocol with core questions and inclusion criteria before commencement. There were no cases in which the clear purpose of the study was not provided.

(2) Was there duplicate study selection and data extraction?

The SR or MA is conducted independently by at least two researchers during the processes of study selection and data extraction. The SR or MA suggests a consensus process for the resolution of disagreements. Among the evaluated SRs and MAs, 81.3% (183/225 studies) were conducted by at least two or more researchers.

(3) Was a comprehensive literature search performed?

SRs and MAs must employ literature searches of at least two or more electronic sources. However, it was found that two or more electronic sources were used in 90.7% (204/225) of the studies, which means that 9% of the studies were conducted with only one database or no electronic sources.

(4) Were the inclusion/exclusion criteria reported clearly?

This question is used to judge the literature searches (i.e., whether the studies are published, whether the studies are extracted according to publication status, and language). If grey

Table 3. Analysis of Methodological Quality

Methodological quality the ratio of "yes"	No. of studies (%)				Total
	2011	2012	2013	Sep 2014	
80%–100% (++++)	21 (41.2)	33 (66.0)	50 (74.6)	33 (57.9)	137 (60.9)
60%–80% (+++)	16 (31.3)	13 (26.0)	12 (17.9)	12 (21.1)	53 (23.5)
40%–60% (++)	9 (17.7)	3 (6.0)	2 (3.0)	10 (17.5)	24 (10.7)
<40% (+)	5 (9.8)	1 (2.0)	3 (4.5)	2 (3.5)	11 (4.9)
Total	51 (100.0)	50 (100.0)	67 (100.0)	57 (100.0)	225 (100.0)

literature and unpublished literature are considered, this question could be answered as "yes." However, when performing the actual research, it is very difficult to extract and compare data that are based on the results of grey literature and unpublished literature. Therefore, we used a revised question for our evaluation: "Was the inclusion/exclusion criteria reported?" The conclusion was that the selection or exclusion of criteria was reported in 91.1% (205/225) of studies.

(5) Was a list of studies (i.e., included studies) provided?

The SRs and MAs that are presented in the bibliography are the included and excluded studies. However, it is difficult to present all of the excluded studies in a journal with page limits. Thus, as long as the selected studies were presented in the bibliography, this question was answered as "yes." According to this standard, it was found that 94.7% (213/225) of the papers listed all included studies in the reference.

(6) Were the characteristics of the included studies provided?

The SRs and MAs should include the subject, intervention, outcome, and characteristics of the study (e.g., age, race, gender, disease state, duration, severity, combined disease) in an organized manner. We found that the characteristics of studies were provided in 90.7% (204/225) of the papers.

(7) Was the scientific quality of the included studies assessed and documented?

SRs and MAs should include analyses of sensitivity, including the risk of bias, and quality. However, the risk of bias and the quality of the appropriate tools and checklists were evaluated in only 57.8% (130/225) of the SAs and MAs.

(8) Was the scientific quality of the included studies appropriately considered in the formulation of conclusions?

During data analysis and the drawing of conclusions, the evaluation of the quality and the risk of bias of the included studies should be considered. However, the results of the methodological rigidity analysis and the quality evaluation were only considered in 56.4% (127/225) of the studies.

(9) Were the methods that were used to combine the findings of studies appropriate?

The combined possibility was verified by evaluating the homogeneity of each study. If there is heterogeneity, proper statistical methods should be used. This question does not refer to SRs. All but 4.4% of the analyses were of high quality and were

conducted appropriately.

(10) Was the likelihood of publication bias assessed?

In SRs and MAs, publication bias should be considered. In the present investigation, we found that an assessment of publication bias by funnel plot analysis or other statistical methods was performed in only 54.2% (142/225) of the studies.

(11) Was the conflict of interest stated?

The funding sources and the conflict of interest were reported in 80.9% of all studies (182/225 studies). The following results were based on the percentage of "yes" responses to each of the following items that were summarized above: ++++, the percentage of "yes" responses to 80%–100% of the items (nine or more items) was observed in 60.9% (137/225) of the studies; +++, the percentage of "yes" responses to 60% to 80% (seven to eight) of the items was observed in 23.6% (53/225) of the studies; ++, the percentage of "yes" responses to 40% to 60% (five to six) of the items was observed in 11.1% (24/225) of the studies; and +, the percentage of "yes" responses to less than 40% of the items (four items or less) was observed in 4.9% (11/225) of the studies (Table 3).

A summary of the overall evaluation of the quality of SRs and MAs using AMSTAR is as follows. The items that received a good evaluation score were the following: "Was the research question (research purpose) clarified?", "Were the inclusion/exclusion criteria reported clearly?", "Was a list of studies (included studies) provided?", "Was a comprehensive literature search performed?", "Were the characteristics of the included studies provided?", and "Were the methods used to combine the findings of studies appropriate?" In contrast, the items that received a poor score were the following: "Was the scientific quality of the included studies assessed and documented?", "Was the scientific quality of the included studies used appropriately in the formulation of conclusions?", and "Was the likelihood of publication bias assessed?"

CONCLUSIONS

We examined the quality and quantity of SRs and MAs related to the field of hepatology that were published in the SCI and SCIE by using the AMSTAR measurement tool. The quality of the results for the 11 criteria assessed by the AMSTAR was

found to be generally high. However, the results associated with the following three criteria were found to be of low quality: (1) lack of scientific quality in the assessment and documentation of the included studies; (2) lack of consideration of the scientific quality of the included studies in the formulation of conclusions; (3) lack of consideration of the likelihood of publication bias.

To draw quality conclusions from the results of the included studies, SRs and MAs need to be conducted using a stricter and more objective research process with the cooperation of clinical experts and methodological professionals. In this manner, bias and random errors can be minimized.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

REFERENCES

1. Kim G, Baik SK. Overview and recent trends of systematic reviews and meta-analyses in hepatology. Clin Mol Hepatol 2014;20:137-150.
2. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [Internet]. London: The Cochrane Collaboration; 2011 [cited 2014 Oct 15]. Available from: http://handbook.cochrane.org.
3. Twaddle S, Qureshi S. Scottish intercollegiate guidelines network. Evid Based Healthc Public Health 2005;9:405-409.
4. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264-269.
5. 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;62:1013-1020.
6. 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.