Low Methodological Quality of Systematic Reviews on Acupuncture: A Cross- Sectional Study

Leonard TF Ho
The Chinese University of Hong Kong
Fiona YT Ke
The Chinese University of Hong Kong
Charlene HL Wong
The Chinese University of Hong Kong
Irene XY Wu (irenexywu@csu.edu.cn)
Central South University
Andy KL Cheung
The Chinese University of Hong Kong
Chen Mao
Southern Medical University
Vincent CH Chung
The Chinese University of Hong Kong

Research Article

Keywords: Evidence-based practice, Meta-analysis, acupuncture, Research design, Systematic reviews

Posted Date: August 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-778930/v1

Abstract

Background: While well-conducted systematic reviews (SRs) can provide best evidence on the potential effectiveness of acupuncture, limitations on methodological rigour of SRs may impact trustworthiness of their conclusions. This cross-sectional study aimed to evaluate the methodological quality of a representative sample of SRs on acupuncture effectiveness.

Methods: CDSR, MEDLINE, and EMBASE were searched for SRs on acupuncture. AMSTAR2 was applied for assessing methodological quality. Associations between bibliographical characteristics and methodological quality ratings were examined.

Results: A total of 106 SRs were appraised. Only one (0.9%) SR was of high overall methodological quality, zero (0%) was of moderate-quality, six (5.7%) and 99 (93.4%) were of low-quality and critically low-quality respectively. Among appraised SRs, only ten (9.4%) provided an *a priori* protocol, four (3.8%) conducted a comprehensive literature search, five (4.7%) provided a list of excluded study, and six (5.7%) performed meta-analysis appropriately. Cochrane reviews, update reviews, reviews with corresponding authors from the America, and reviews that searched non-English databases had relatively higher overall quality.

Conclusions: Methodological quality of SRs on acupuncture is unsatisfactory. Future reviewers should improve critical areas of publishing protocols, performing comprehensive search, providing a list of excluded studies with justifications for exclusion, and conducting meta-analysis appropriately.

Introduction

The delivery of traditional, complementary, and integrative medicine (TCIM) services in an evidence-based manner is advocated by the World Health Organization (WHO) in its *Traditional Medicine Strategy 2014–2023*.[1] As a popular form of TCIM, the use of acupuncture is increasing globally. In China, traditional Chinese medicine (TCM) constitutes a formal part of the health system, of which 20% of all outpatient services were delivered by the TCM sector, including acupuncture.[2] In Taiwan, acupuncture services are covered by the National Health Insurance,[3] and the prevalence of acupuncture use was 11% in 2011.[4] Meanwhile, in Australia,[5] Germany,[6] and Norway,[7] acupuncture is not only regulated by the government or relevant authorities but is also partially or fully covered by statutory health insurance.

In response to the WHO's initiative, there is a need to synthesise up-to-date evidence on the effectiveness of acupuncture, so as to facilitate the implementation of evidence-based acupuncture services. Although the number of systematic reviews (SRs) on acupuncture effectiveness has been increasing recently,[8] there are still concerns over their methodological quality.[9, 10] For instance, inappropriate literature search, absence of critical appraisal of included primary studies, and meta-analysis of highly heterogeneous studies may give rise to biased conclusions.[11, 12]

This cross-sectional study aimed to (i) describe the bibliographical characteristics of SRs on acupuncture trials; (ii) appraise the methodological quality of SRs on acupuncture trials using AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews 2) instrument.[10]

Methods

Eligibility criteria

Acupuncture refers to the use of stainless-steel filiform needles to puncture specific acupoints on the body to trigger specific therapeutic effects.[13] SRs published in English with at least one meta-analysis on the treatment effect of acupuncture, including traditional manual acupuncture and electro-acupuncture, were eligible. SRs on acupuncture with moxibustion, a TCM therapy involving the burning of herbs over the skin,[14] were also included. SRs on transcutaneous electrical nerve stimulation and laser acupuncture were excluded. Animal studies, narrative reviews, protocol, and network meta-analyses were also ineligible. For duplicates of SRs, the most updated versions were included for appraisal.

Literature search

A comprehensive literature search was conducted in three international electronic databases, including the Cochrane Database of Systematic Reviews, MEDLINE, and EMBASE, for a representative sample of SRs published from January 2018 to March 2020. It is recommended that SR should be updated every two years, and hence we have chosen a sampling time frame that allowed us to focus on current SRs.[9] Details on the search strategies are shown in eTable 1, Additional file 1. Validated search filters for SRs were applied to maximise the specificity of search on MEDLINE [15] and EMBASE.[16]

Literature screening and data extraction

All retrieved citations were imported into Endnote X9. After deduplication, titles and abstracts of retrieved citations were screened against the eligibility criteria. Full texts of potentially eligible citations were subsequently retrieved for further assessment. For included SRs, bibliographic characteristics were extracted using a pre-designed questionnaire (eTable 2, Additional file 1).[17–20]

Literature selection and data extraction were conducted by two independent reviewers (FYTK and AKLC). Disagreements and discrepancies were resolved via consensus between reviewers, or by arbitration of a third reviewer (CHLW).

Methodological quality assessment

Methodological quality of included SRs was appraised by the validated AMSTAR 2 instrument,[10] across all 16 domains. Seven domains were considered as critical:

1. Protocol registered before commencement of the review (item 2)
2. Adequacy of the literature search (item 4)

3. Justification for excluding individual studies (item 7)
4. Risk of bias from individual studies being included in the review (item 9)
5. Appropriateness of meta-analytical methods (item 11)
6. Consideration of risk of bias when interpreting the results of the review (item 13)
7. Assessment of presence and likely impact of publication bias (item 15)

Based on their performance on each domain, each SRs were rated as being “high”, “moderate”, “low”, and “critically low” in terms of overall methodological quality,[10] in accordance with published operational guidelines (eTable 3, Additional file 1). Methodological quality assessment was conducted by two authors (FYTK and LTFH) independently. Disagreements and discrepancies were resolved via consensus between authors, or by arbitration of a senior researcher (VCHC).

Data analysis

Data on bibliographical characteristics and AMSTAR 2 methodological quality assessment results were summarised using descriptive statistics. Differences in the overall methodological quality of SRs across different bibliographical characteristics were investigated using Kruskal-Wallis rank tests and Spearman's rank correlation coefficients. A p-value < 0.05 was considered statistically significant. All statistical analyses were performed using SPSS 26.

Results

Literature selection

The literature search yielded a total of 1,065 citations. After deduplication, titles and abstracts of 764 citations were screened. Then, 185 publications proceeded to full-text eligibility assessment. Finally, 106 SRs fulfilled the eligibility criteria and were included (eTable 4, Additional file 1). Details on literature selection are illustrated in Fig. 1. A full list of excluded records is presented in eTable 5, Additional file 1.

Bibliographical characteristics of the included systematic reviews

The 106 included SRs contained 1,864 randomised controlled trials with 204,784 participants. Only five SRs (4.7%) were Cochrane reviews. Nineteen SRs (17.9%) were an update of previous SRs. The median publication year was 2019. Impact factor ranged from 0 to 6.8 with a median of 2.0. Number of review authors ranged from two to 13 with a median of six. The corresponding authors of 93 (87.7%) SRs were from Asia, seven (6.6%) from America, four (3.8%) from Europe, and two (1.9%) from Oceania. Over a half (66; 62.3%) of the SRs had their funding sources located in Asia, while 24 (22.6%) SRs did not receive any funding support.

One-hundred-and-five (99.1%) SRs involved English database searching, while 88 (83.0%) involved non-English database searching. Most SRs reported both starting and ending years of search (81; 76.4%) and

search terms for one or more electronics databases (101; 95.2%). Seventy-six (71.7%) SRs reported intervention harms. Nevertheless, 59 (55.7%) SRs did not report the language of the included primary studies.

Ninety-nine (93.4%) applied the *Cochrane risk-of-bias tool* for assessing risk of bias, two used *Jadad scale* (1.9%) or *Pedro scale* (1.9%) respectively, and two did not perform risk of bias assessment (1.9%). One-hundred-and-two (96.2%) SRs included a PRISMA-like flow diagram to illustrate the process of literature selection. Details on bibliographical characteristics are shown in Table 1.

Table 1
Bibliographical characteristics of the 106 included systematic reviews on acupuncture

Bibliographical characteristics	Results*
Cochrane review	5 (4.7)
An update of previous review	19 (17.9)
An update of previous Cochrane review	3 (2.8)
An update of a previous non-Cochrane review	16 (15.1)
Publication year median (range)	2019 (2018–2020)
Publication journal impact factor median (range)	2.0 (0-6.8)
Number of review authors median (range)	6 (2–13)
Location of corresponding author	
Europe	4 (3.8)
America	7 (6.6)
Asia	93 (87.7)
Oceania	2 (1.9)
Number of included primary studies	
Total	1864
Median (range)	13.5 (3–73)
Number of participants included in primary studies	
Total	204784
Median (range)	1238 (178–20827)
SRs reporting intervention harms	76 (71.7)
Funding location of the SR	
Europe	4 (3.8)
America	4 (3.8)
Asia	66 (62.3)
Not reported	8 (7.5)

MeSH: Medicine Medical Subject Headings; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; SR: Systematic review.

*Values are *n* (%) or median (range).

Bibliographical characteristics	Results*
No funding support	24 (22.6)
SRs that searched English databases	105 (99.1)
SRs that searched non-English databases	88 (83.0)
Report year span of search	
Yes, reported both starting and ending years	81 (76.4)
Partially, only reported starting years	19 (17.9)
Not mentioned	6 (5.7)
Search terms reported for one or more electronic databases	
Topics/free text/keywords/MeSH	47 (44.3)
Full Boolean	54 (50.9)
Readers are referred elsewhere for full search strategy	0 (0)
No research term	5 (4.7)
Language of included primary studies in SRs	
English only	9 (8.5)
Language other than English	6 (5.7)
English and other languages	32 (30.2)
Not reported	59 (55.7)
Risk of bias assessment tools	
Cochrane risk of bias	99 (93.4)
Jadad scale	2 (1.9)
Pedro Scale	2 (1.9)
Others	1 (0.9)
Risk of bias assessment tool not used	2 (1.9)
Included a PRISMA-like flow diagram	102 (96.2)

MeSH: Medicine Medical Subject Headings; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; SR: Systematic review.

*Values are *n* (%) or median (range).

Methodological quality

Performance among the included SRs was inadequate across four of the seven AMSTAR 2 critical domains, with that less than 20% satisfying the following: ten (9.4%) SRs established an *a priori* protocol and justified deviations from the protocol (item 2); four (3.8%) implemented a comprehensive literature search strategy (item 4); five (4.7%) listed excluded studies and justified the exclusions (item 7); and six (5.7%) conducted appropriate meta-analysis (item 11).

Included SRs performed relatively better across the remaining three critical domains: 97 (91.5%) had the risk of bias of individual studies assessed by appropriate instruments (item 9); 78 (73.6%) accounted for risk of bias among individual studies when interpreting results (item 13); and 23 (21.7%) investigated publication bias, and discussed its potential impact on the results (item 15).

Performance was unsatisfactory among four of the nine non-critical domains, with less than 20% fulfilling relevant criteria: four (3.8%) explained the selection of study designs for inclusion (item 3); 12 (11.3%) described included studies in adequate details (item 8); four (3.8%) reported sources of funding among individual studies included (item 10); and 19 (17.9%) assessed potential impact of risk of bias among individual studies on the results of meta-analysis (item 12).

More than 75% of SRs performed well across the remaining five non-critical domains: all SRs reported the PICO components (Problem/Patient/Population, Intervention/Indicator, Comparison, and Outcome) in their research questions and inclusion criteria (item 1); 94 (88.7%) and 97 (91.5%) SRs performed study selection (item 5) and data extraction (item 6) in duplicate, respectively; 84 (79.2%) provided a satisfactory explanation for heterogeneity in the results (item 14); and nearly all (104; 98.1%) reported the potential sources of conflict of interest (item 16). Details on the results of the AMSTAR 2 items are illustrated in Table 2.

Table 2
Results of the AMSTAR-2 items for the 106 systematic reviews on acupuncture

AMSTAR- 2 items	Yes (%)	Partial Yes (%)	No (%)
1. Did the research questions and inclusion criteria for the review include the components of PICO?	106 (100)	NA	0 (0)
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?*	10 (9.4)	32 (30.2)	64 (60.4)
3. Did the review authors explain their selection of the study designs for inclusion in the review?	4 (3.8)	NA	102 (96.2)
4. Did the review authors use a comprehensive literature search strategy? *	4 (3.8)	99 (93.4)	3 (2.8)
5. Did the review authors perform study selection in duplicate?	94 (88.7)	NA	12 (11.3)
6. Did the review authors perform data extraction in duplicate?	97 (91.5)	NA	9 (8.5)
7. Did the review authors provide a list of excluded studies and justify the exclusions?*	5 (4.7)	1 (0.9)	100 (94.3)
8. Did the review authors describe the included studies in adequate detail?	12 (11.3)	84 (79.2)	10 (9.4)
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?*	97 (91.5)	3 (2.8)	6 (5.7)
10. Did the review authors report on the sources of funding for the studies included in the review?	4 (3.8)	NA	102 (96.2)
11. If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?*	6 (5.7)	NA	100 (94.3)
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	19 (17.9)	NA	87 (82.1)
13. Did the review authors account for RoB in individual studies when interpreting / discussing the results of the review?*	78 (73.6)	NA	28 (26.4)
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	84 (79.2)	NA	22 (20.8)

AMSTAR 2: A MeaSurement Tool to Assess systematic Reviews 2; NA: Not applicable.

*Critical domain-specific item.

AMSTAR- 2 items	Yes (%)	Partial Yes (%)	No (%)
15. If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?*	23 (21.7)	NA	83 (78.3)
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	104 (98.1)	NA	2 (1.9)
AMSTAR 2: A MeaSurement Tool to Assess systematic Reviews 2; NA: Not applicable.			
*Critical domain-specific item.			

Relationship between bibliographical characteristics and overall methodological quality

Among the 106 appraised SRs, only one (0.9%) of them was of high overall methodological quality, while six (5.7%) were of low-quality. The remaining 99 (93.4%) SRs were of critically low-quality.

Results of Kruskal-Wallis tests indicated that there were statistically significant between-group differences across five bibliographical characteristics [Table 3]. Cochrane reviews ($p < 0.001$), an update of a previous non-Cochrane review ($p = 0.007$), SRs published in 2018 ($p = 0.014$), SRs with corresponding authors from America ($p = 0.028$), and SRs that searched non-English databases ($p = 0.048$) had higher overall methodological quality than their counterparts. The Spearman's rank correlation coefficient also showed that SRs published in higher impact factor journals ($r_s = 0.36$; $p < 0.001$) were associated with higher overall methodological quality.

Table 3
Overall methodological quality of the 106 systematic reviews on acupuncture by bibliographical characteristics

Bibliographical characteristics	Critically low^	Low^	Moderate^	High^	*P*
Total included SRs	99 (93.4)	6 (5.7)	0 (0.0)	1 (0.9)	
Cochrane Review					< 0.001*
Yes	0 (0.0)	4 (80.0)	0 (0.0)	1 (20.0)	
No	99 (98.0)	2 (2.0)	0 (0.0)	0 (0)	
An update of a previous review					0.007*
Yes (Cochrane review)	3 (100)	0 (0)	0 (0.0)	0 (0.0)	
Yes (non-Cochrane review)	12 (75.0)	3 (18.8)	0 (0.0)	1 (6.3)	
No	84 (96.6)	3 (3.4)	0 (0.0)	0 (0.0)	
Published year					0.014*
2018	40 (85.1)	6 (12.8)	0 (0.0)	1 (2.1)	
2019	49 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
2020	10 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Location of corresponding author					0.028*
Europe	4 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
America	6 (85.7)	0 (0.0)	0 (0.0)	1 (14.3)	

MeSH: National Library of Medicine Medical Subject Headings; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; SR: Systematic review.

^Values are *n* (% in subgroup).

*P value of Kruskal-Wallis test was < 0.05.

Bibliographical characteristics	Critically low[^]	Low[^]	Moderate[^]	High[^]	*P*
Asia	87 (93.5)	6 (6.5)	0 (0.0)	0 (0.0)	
Oceania	2 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Reported intervention harms					0.659
Yes	70 (92.1)	5 (6.6)	0 (0.0)	1 (1.3)	
No	29 (96.7)	1 (3.3)	0 (0.0)	0 (0.0)	
Funding location of the SR					0.859
Europe	4 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
America	4 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Asia	61 (92.4)	4 (6.1)	0 (0.0)	1 (1.5)	
Not reported	7 (87.5)	1 (12.5)	0 (0.0)	0 (0.0)	
No funding support	23 (95.8)	1 (4.2)	0 (0.0)	0 (0.0)	
SRs that searched non-English databases					0.048[*]
Yes	82 (93.2)	6 (6.8)	0 (0.0)	0 (0.0)	
No	17 (94.4)	0 (0.0)	0 (0.0)	1 (5.6)	
Report year of coverage of literature search					0.323
Yes	74 (91.4)	6 (7.4)	0 (0.0)	1 (1.2)	
Partially	19 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	

MeSH: National Library of Medicine Medical Subject Headings; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; SR: Systematic review.

[^]Values are *n* (% in subgroup).

[*]P value of Kruskal-Wallis test was < 0.05.

Bibliographical characteristics	Critically low^	Low^	Moderate^	High^	*P*
Not mentioned	6 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Search terms reported for one or more electronic databases					0.287
Topics/free text/keywords/MeSH	47 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Full Boolean	48 (88.9)	5 (9.3)	0 (0.0)	1 (1.9)	
Readers are referred elsewhere for full search strategy	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
No research term	4 (80.0)	1 (20.0)	0 (0.0)	0 (0.0)	
Eligibility criteria based on language of publication					0.467
English only	9 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Language other than English	5 (83.3)	1 (16.7)	0 (0.0)	0 (0.0)	
English and other languages	28 (87.5)	3 (9.4)	0 (0.0)	1 (3.1)	
Not reported	57 (96.6)	2 (3.4)	0 (0.0)	0 (0.0)	
Risk-of-bias assessment tools					0.769
Cochrane risk of bias	92 (92.9)	6 (6.1)	0 (0.0)	1 (1.0)	
Jadad scale	2 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Pedro Scale	2 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Others	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Not mentioned	2 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	

MeSH: National Library of Medicine Medical Subject Headings; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; SR: Systematic review.

^Values are *n* (% in subgroup).

*P value of Kruskal-Wallis test was < 0.05.

Bibliographical characteristics	Critically low^	Low^	Moderate^	High^	*P*
Included a PRISMA-like flow diagram					0.865
Yes	95 (93.1)	6 (5.9)	0 (0.0)	1 (1.0)	
No	4 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
MeSH: National Library of Medicine Medical Subject Headings; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; SR: Systematic review.					
^Values are *n* (% in subgroup).					
*P value of Kruskal-Wallis test was < 0.05.					

Discussion

Summary of results

This cross-sectional study investigated the methodological quality of a representative sample of 106 SRs on acupuncture effectiveness published between 2018 to 2020. Our results revealed that the methodological rigour of acupuncture SRs is weak, with more than 93% being at critically low-quality. Being Cochrane review, an update of a previous non-Cochrane review, SRs with corresponding authors from America, SRs that searched non-English databases, and being published in journals with higher impact factor are associated with better quality, but they only constitute a small number of SRs.

Recommendation for future systematic reviews

Publishing an a priori review protocol

As SR authors tend to include primary studies with positive results,[21] publication of an *a priori* SR protocol would reduce selective outcome reporting and enable comparison of SR protocol and its publications.[9, 22] This also minimises influence of reviewers' biases caused by foreknowledge on preliminary results, allows peer-reviewing of planned methods, and reduces research waste due to duplication.[9] Our study showed that only 9.4% of SRs satisfied this criterion. Future authors should publish SR protocols in open-access journals, or register them on international databases,[9, 10] like the International Prospective Register of Systematic Reviews (PROSPERO).[23]

Conducting comprehensive literature search

Only 3.8% fulfilled the criteria for completing a comprehensive literature search, as many did not conduct searches on trial registries, conference abstracts, theses, and grey literature, examining reference lists of included studies, and consulting the experts in the field of acupuncture. Such incomprehensiveness may

give rise to publication bias, leading to over-estimation of effectiveness.[24, 25] On average, the exclusion of grey literature may result in an overestimation of intervention effect by approximately 12%.[26] It is also noteworthy that 55.7% did not report the language of included primary studies, which cast doubts on whether non-English publications were included. If this is the case, language bias may occur,[9] leading to an overestimation or underestimation of intervention effect.[27, 28] For future acupuncture SR, ensuring a search for grey as well as non-English literature is a clear area for future improvement.

Providing a list of excluded studies and detailed description of included primary studies

A list of excluded studies with justifications for exclusion promotes transparency and reproducibility of SRs.[29, 30] Such a list may reduce potential publication bias and exclusion errors caused by unavoidable subjectivity during the study selection process.[9, 10] On the other hand, for included primary studies extensive details on PICO elements should be reported. These details can assist evidence users in evaluating the external validity and applicability of the findings,[9] as well as in facilitating the exploration of clinical heterogeneity across primary studies.[9, 10, 31] Unfortunately, only 4.7% and 11.3% of SRs provided a list of excluded studies with rationales, or described the included primary studies in detail, respectively. Future reviewers should avoid these limitations.

Conducting appropriate meta-analysis

Our findings indicated that 94.3% applied inappropriate meta-analysis methods, mainly due to improper choice for a fixed-effect model. This model assumes that there is only one true effect size among all included studies, and the pooled effect estimate is common to all studies. On the contrary, a random-effect model assumes that the true effect size varies among studies, and pooled effect estimate is the mean of a distribution of true effects.[32] In the context of a meta-analysis of acupuncture trials, a random effect model is the correct choice in most of the cases. This is because some heterogeneity across trials is expected, and the assumption that there is only one true effect size is unlikely to hold.[9, 32]

Strengths and limitations

This study applied AMSTAR 2 to evaluate an up-to-date representative sample of SRs on acupuncture effectiveness. While our results indicated an urgent need to improve methodological quality of SRs in the field, there are also several limitations. This cross-sectional study only appraised SRs published in English, which may limit the generalisability of our results. That said, it is expected that SRs published in English remain to be a key source of evidence impacting clinical decision on acupuncture internationally. Also, since quality assessment of this study merely relied on published information, poor adherence to reporting guidelines and word limit of the journal may indeed affect the accuracy of assessment.[8]

Implications

Most SRs on acupuncture effectiveness are of critically low methodological quality, of which these may give rise to an underestimation or overestimation of treatment effectiveness. Healthcare providers,

guideline developers, and other evidence users should critically appraise the methodological quality of SRs before applying relevant evidence in policy- and clinical decision-making. Journal editors and peer-reviewers are also recommended to use AMSTAR 2 [10] and *Cochrane Handbook for Systematic Reviews of Interventions* [9] as guidelines for verifying quality of SRs submitted.

Internationally, recommendations on the use of acupuncture have been increasing in clinical practice guidelines.[33, 34] In recent years, the Chinese Government has been upscaling the resources allocated to acupuncture research and development, as well as advocating evidence-based acupuncture practice.[35] Unfortunately, findings of this study revealed that the methodological rigour of SRs on acupuncture may not be robust enough to support these guidelines and policy initiatives. Prior to extensive implementation of evidence-based acupuncture services, clinical epidemiology education, particularly on performing high-quality clinical research and synthesis, should be emphasised.

Conclusions

Methodological quality of SRs on acupuncture published in recent years was unsatisfactory, with only 0.9% of them being of high overall quality. For future SR authors, improvement efforts should focus on publishing *a priori* research protocols, conducting comprehensive literature search, providing lists of excluded studies with justifications for exclusion, and employing appropriate methods for meta-analysis.

Abbreviations

AMSTAR 2: A MeaSurement Tool to Assess systematic Reviews 2; PROSPERO: International Prospective Register of Systematic Reviews; SR: Systematic review; TCM: Traditional Chinese medicine; TCIM: Traditional, complementary, and integrative medicine; WHO: World Health Organization.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

This work was supported by the National Natural Science Foundation of China (81973709) and the Hunan Nature Science Foundation (2019JJ40348).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LTFH: Collection and assembly of data; Data analysis and interpretation; Manuscript writing. FYTK: Collection and assembly of data; Data analysis and interpretation. CHLW: Data analysis and interpretation; Manuscript writing. IXYW: Conception and design; Resources; Funding acquisition. AKLC: Data analysis and interpretation. MC: Conception and design. VCHC: Conception and design; Manuscript writing; Supervision. The authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. World Health Organization. WHO Traditional Medicine Strategy: 2014–2023. Geneva: World Health Organization; 2013.
2. Chung VCH, Ma PHX, Wang HHX, Wang JJ, Hong LC, Wei X, et al. Integrating Traditional Chinese Medicine Services in Community Health Centers: Insights into Utilization Patterns in the Pearl River Region of China. Evid Based Complement Alternat Med. 2013;2013:426360.
3. Chen FP, Kung YY, Chen TJ, Hwang SJ. Demographics and patterns of acupuncture use in the Chinese population: the Taiwan experience. J Altern Complement Med. 2006;12(4):379–87.
4. Wu MY, Lee YC, Lin CL, Huang MC, Sun MF, Yen HR. Trends in use of acupuncture among adults in Taiwan from 2002 to 2011: A nationwide population-based study. PLoS ONE. 2018;13(4):e0195490.
5. Zheng Z. Acupuncture in Australia: regulation, education, practice, and research. Integr Med Res. 2014;3(3):103–10.
6. Trinczek K. Reimbursement for Acupuncture Treatments in the German Statutory Health Insurance System. Forsch Komplementmed. 2015;22(2):118–23.
7. Jacobsen R, Fønnebø VM, Foss N, Kristoffersen AE. Use of complementary and alternative medicine within Norwegian hospitals. BMC Complement Altern Med. 2015;15(1):275.
8. Liu Y, Zhang R, Huang J, Zhao X, Liu D, Sun W, et al. Reporting Quality of Systematic Reviews/Meta-Analyses of Acupuncture. PLoS ONE. 2014;9(11):e113172.
9. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. London, United Kingdom: The Cochrane Collaboration. Accessed Jan 4 2021. Available from: https://training.cochrane.org/handbook.

10. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.
11. 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 Med Res Methodol. 2007;7(1):10.
12. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008;358(3):252–60.
13. Zhang J, Shang H, Gao X, Ernst E. Acupuncture-related adverse events: a systematic review of the Chinese literature. Bull World Health Organ. 2010;88(12):915-21c.
14. World Health Organization. Acupuncture: Review and Analysis of Reports on Controlled Clinical Trials. Geneva: World Health Organization; 2003.
15. McMaster Health Information Research Unit. Search Filters for MEDLINE in Ovid Syntax and the PubMed translation. Hamilton, Ontario: McMaster Health Information Research Unit. Accessed Jan 4 2021. Available from: https://hiru.mcmaster.ca/hiru/HIRU_Hedges_MEDLINE_Strategies.aspx.
16. McMaster Health Information Research Unit. Search Strategies for EMBASE in Ovid Syntax. Hamilton, Ontario: McMaster Health Information Research Unit. Accessed Jan 4 2021. Available from: https://hiru.mcmaster.ca/hiru/HIRU_Hedges_EMBASE_Strategies.aspx.
17. Wu XY, Lam VCK, Yu YF, Ho RST, Feng Y, Wong CHL, et al. Epidemiological characteristics and methodological quality of meta-analyses on diabetes mellitus treatment: a systematic review. Eur J Endocrinol. 2016;175(5):353.
18. Wu XY, Du XJ, Ho RS, Lee CC, Yip BH, Wong MC, et al. Characteristics and Methodological Quality of Meta-Analyses on Hypertension Treatments-A Cross-Sectional Study. J Clin Hypertens (Greenwich). 2017;19(2):137–42.
19. Ho RST, Wu X, Yuan J, Liu S, Lai X, Wong SYS, et al. Methodological quality of meta-analyses on treatments for chronic obstructive pulmonary disease: a cross-sectional study using the AMSTAR (Assessing the Methodological Quality of Systematic Reviews) tool. NPJ Prim Care Respir Med. 2015;25(1):14102.
20. Tsoi AKN, Ho LTF, Wu IXY, Wong CHL, Ho RST, Lim JYY, et al. Methodological quality of systematic reviews on treatments for osteoporosis: A cross-sectional study. Bone. 2020;139:115541.
21. Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. JAMA. 2002;287(21):2831–4.
22. The PLoS Medicine Editors. Best practice in systematic reviews: the importance of protocols and registration. PLoS Med. 2011;8(2):e1001009.
23. Centre for Reviews and Dissemination - University of York. About PROSPERO York, United Kingdom: Centre for Reviews and Dissemination - University of York. Accessed Jan 4 2021. Available from: https://www.crd.york.ac.uk/prospero/#aboutpage.

24. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ. 2012;344:d7202.
25. Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. Cochrane Database Syst Rev. 2014;2014(4):Cd008965.
26. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet. 2000;356(9237):1228–31.
27. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012;28(2):138–44.
28. Wu XY, Tang JL, Mao C, Yuan JQ, Qin Y, Chung VC. Systematic Reviews and Meta-Analyses of Traditional Chinese Medicine Must Search Chinese Databases to Reduce Language Bias. Evid Based Complement Alternat Med. 2013;2013:812179.
29. Faggion CM. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. BMC Med Res Methodol. 2015;15(1):63.
30. Wegewitz U, Weikert B, Fishta A, Jacobs A, Pieper D. Resuming the discussion of AMSTAR: What can (should) be made better? BMC Med Res Methodol. 2016;16(1):111.
31. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ. 1994;309(6965):1351–5.
32. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97–111.
33. Birch S, Lee MS, Alraek T, Kim TH. Overview of Treatment Guidelines and Clinical Practical Guidelines That Recommend the Use of Acupuncture: A Bibliometric Analysis. J Altern Complement Med. 2018;24(8):752–69.
34. Guo Y, Zhao H, Wang F, Li S-n, Sun Y-x, Han M-j, et al. Recommendations for acupuncture in clinical practice guidelines of the national guideline clearinghouse. Chin J Integr Med. 2017;23(11):864–70.
35. The State Council of the People's Republic of China. Opinions of the CPC Central Committee and the State Council on Promoting the Preservation, Innovation, and Development of Traditional Chinese Medicine (Chinese). China: The State Council of the People's Republic of China. Accessed Jan 4 2021. Available from: http://www.gov.cn/zhengce/2019-10/26/content_5445336.htm.

Figures

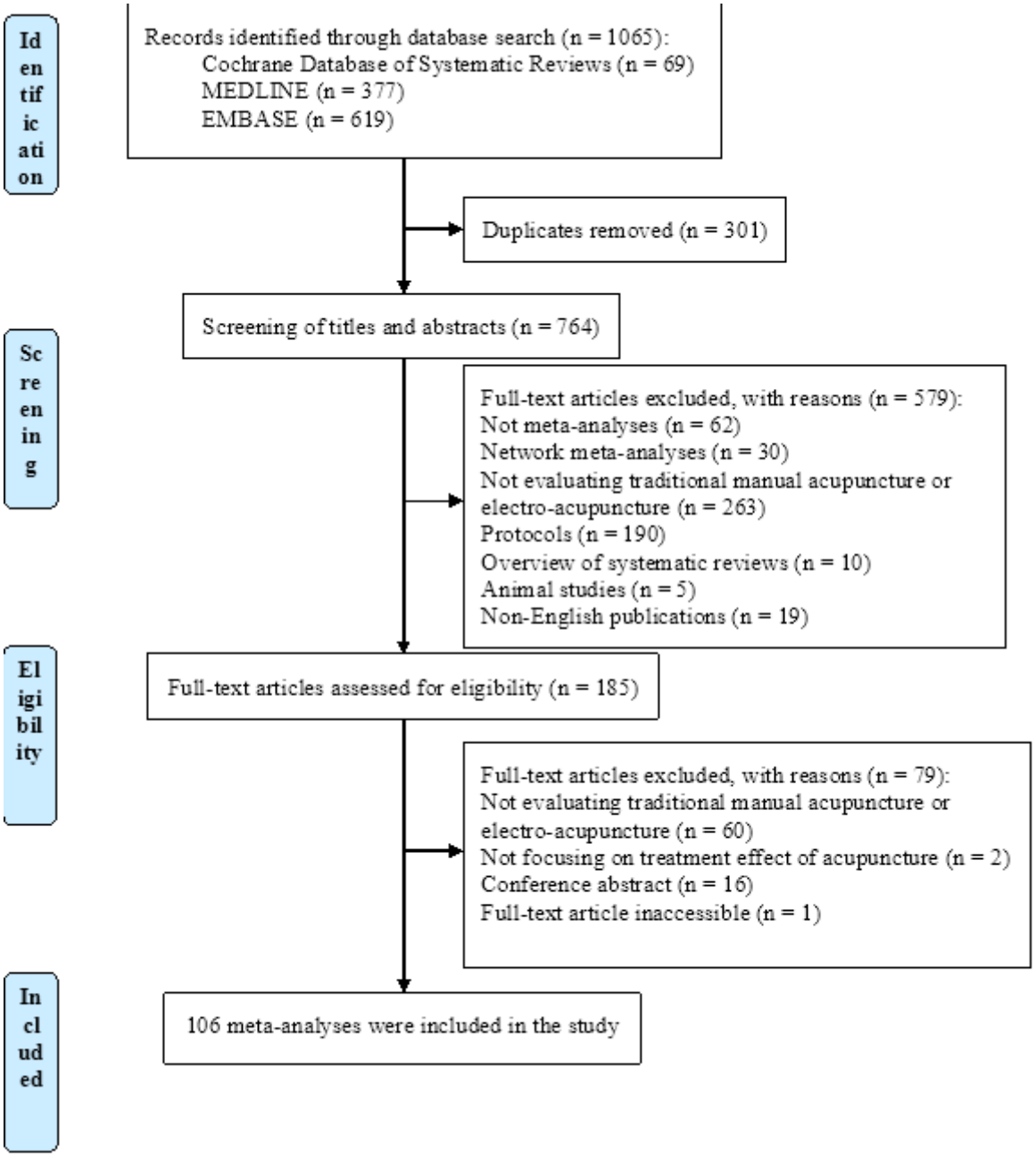

Figure 1

Process of literature selection

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AMSTARAcuadditionalfileMRM.docx