



## ORIGINAL ARTICLE

# The methodological and reporting quality of systematic reviews from China and the USA are similar

Jinhui Tian<sup>a,b</sup>, Jun Zhang<sup>c</sup>, Long Ge<sup>a,b</sup>, Kehu Yang<sup>a,b,\*\*</sup>, Fujian Song<sup>d,\*</sup>

<sup>a</sup>Evidence-based Medicine Center of Lanzhou University, 199 Donggang West Road, Lanzhou City, Gansu, 730000, China

<sup>b</sup>Key Laboratory of Evidence-based Medicine and Knowledge Translation of Gansu Province, 199 Donggang West Road, Lanzhou City, Gansu, 730000, China

<sup>c</sup>School of Nursing, Gansu University of Chinese Medicine, 35 East Dingxi Road, Lanzhou City, Gansu, 730000, China

<sup>d</sup>Department of Population Health and Primary Care, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK

Accepted 9 December 2016; Published online xxxx

## Abstract

**Objectives:** To compare the methodological and reporting quality of systematic reviews by authors from China and those from the United States (USA).

**Study Design and Setting:** From systematic reviews of randomized trials published in 2014 in English, we randomly selected 100 from China and 100 from the USA. The methodological quality was assessed using the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) tool, and reporting quality assessed using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) tool.

**Results:** Compared with systematic reviews from the USA, those from China were more likely to be a meta-analysis, published in low-impact journals, and a non-Cochrane review. The mean summary Assessing the Methodological Quality of Systematic Reviews score was 6.7 (95% confidence interval: 6.5, 7.0) for reviews from China and 6.6 (6.1, 7.1) for reviews from the USA, and the mean summary Preferred Reporting Items for Systematic Reviews and Meta-analyses score was 21.2 (20.7, 21.6) for reviews from China and 20.6 (19.9, 21.3) for reviews from the USA. The differences in summary quality scores between China and the USA were statistically nonsignificant after adjusting for multiple review factors.

**Conclusion:** The overall methodological and reporting quality of systematic reviews by authors from China are similar to those from the USA, although the quality of systematic reviews from both countries could be further improved. © 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Systematic review; Methodological quality; Reporting quality; Risk of bias; Validity; Evidence-based medicine

## 1. Background

Well-conducted systematic reviews and meta-analyses of randomized controlled trials (RCTs) provide the most valid research evidence on effects of health care interventions [1,2]. Systematic review methods (with or without meta-analysis) have been used in medicine and health research since later 1980s in developed countries [3]. The Cochrane Collaboration and other evidence-based health programs have promoted the use of systematic reviewing methods globally [4], including China [5].

It has been anticipated that systematic reviews would help address challenges due to rapid increase in clinical literature [6,7]. However, the successful production of systematic reviews during past decades has raised concerns about whether the exponential increase in published systematic reviews might have actually exacerbated information overload [7–11]. Particularly, the increased production of systematic reviews by authors from China has been considered at least partly responsible for the rapid increase in systematic reviews globally [12,13]. For example, a search in PubMed on January 8, 2016 (see [Supplementary File 1](#) at [www.jclinepi.com](http://www.jclinepi.com) for the search strategy), found that the number of published systematic reviews by authors from China was increased exponentially from only 19 in 2005 to 1,073 in 2014. During the same time period, the production of systematic reviews by authors from the United States (USA) was only moderately increased from 500 in 2005 to 796 in 2014.

Funding: There was no specific funding for this study.

Conflict of interest: None.

\* Corresponding author. Tel.: +44 1603 591253; fax: +44 1603 593752.

\*\* Corresponding author. Tel.: +86 931 8912767; fax: +86 931 8915076.

E-mail addresses: [Fujian.song@uea.ac.uk](mailto:Fujian.song@uea.ac.uk) (F. Song), [yangkh@lzu.edu.cn](mailto:yangkh@lzu.edu.cn) (K. Yang).

<http://dx.doi.org/10.1016/j.jclinepi.2016.12.004>

0895-4356/© 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**What is new?****Key findings**

- The overall methodological and reporting quality of systematic reviews of randomized trials by authors from China were similar to those from the United States (USA). The differences and similarities in specific quality items between China and the USA were identified.

**What this adds to what was known?**

- This is the first study to compare the reporting and methodological quality of systematic reviews of randomized trials by authors from China (a developing country) and the USA (a developed country).

**What is the implication and what should change now?**

- Considering the usefulness of systematic reviews in evidence-based practice and the development of primary research, the systematic reviewing capacity should be strengthened in China. Identified shortcomings in methodological and reporting quality of published systematic reviews should be considered in further training of authors of systematic reviews in the relevant countries.

With the rapid increase in the number of systematic reviews by authors from China, their reporting and methodological quality have been scrutinized in the previous studies [12,14–18]. These studies usually suggested that the reporting and methodological quality of systematic reviews from China were poor and needed to be much improved. However, it is unclear about the quality of systematic reviews by authors from China relative to those from other countries. There was only one previous study that compared meta-analyses of genetic associations by authors from China and those from the USA [10,12]. According to our knowledge, there were no published studies that systematically compared quality of systematic reviews of RCTs of health care interventions by authors from China and those by authors from other countries.

Identification of differences in methodological and reporting quality of systematic reviews by authors from China and developed countries may help appropriately interpret findings from systematic reviews and set priorities in training of systematic reviewers. Specifically, we consider it appropriate to compare systematic reviews by authors from China and those from the USA for the following reasons: Authors from the USA, along with authors from other high-income nations, have been traditionally the main producer of systematic reviews, and a previous study had compared genetic association

meta-analyses by authors from China and the USA [12]. Therefore, the aim of the current study is to compare the main characteristics, methodological, and reporting quality of systematic reviews of health care interventions between China and the USA. Although the reporting quality was assessed, the focus of the current study was on the methodological quality regarding the validity in the process and results of a systematic review.

**2. Methods***2.1. Identification and selection of systematic reviews*

One reviewer (F.S.) searched PubMed on January 8, 2016, to identify relevant systematic reviews (see [Supplementary File 1](#) at [www.jclinepi.com](http://www.jclinepi.com) for the search strategy). Citations of all identified systematic reviews were downloaded to an EndNote database and then exported to a Microsoft Excel spreadsheet. Each of the originally identified records by country was assigned a random number from 0 to 1 (generated by Excel). Then, the records were ordered from the smallest to the largest by assigned random numbers, and the first 100 eligible systematic reviews from each country were selected. If a selected systematic review was not eligible, a successive record was used to replace it until the total number of included systematic reviews was 100 for each country. Included systematic reviews met the following criteria: (1) was a review article and explicitly stated as a systematic review or meta-analysis, with a formal (comprehensive or not) literature search, (2) was fully published in English in 2014, (3) included only RCTs, and (4) had a corresponding author with an affiliation in mainland China or in the USA. We did not formally calculate the number of systematic reviews required because of no information on what would be clinically meaningful differences in quality of systematic reviews between countries.

*2.2. Quality assessment and data extraction*

All authors involved in this study had previous experience of assessing quality of published systematic reviews. Using a data extraction sheet ([Supplementary File 1](#) at [www.jclinepi.com](http://www.jclinepi.com)), one reviewer (I.Z., L.G., or J.H.T.) extracted and a second reviewer (J.H.T. or F.S.) checked data on the main characteristics from included systematic reviews. Any disagreements were resolved by discussion. Data extracted from systematic reviews included the following: the journal in which a systematic review was published, type of systematic reviews (narrative or meta-analysis), the number of authors, countries which coauthors came from, whether the review protocol was registered, diseases of interest, interventions evaluated, primary outcome measures, the number of RCTs included, the number of total participants, and conclusions of the systematic reviews. Impact factors of journals in which systematic

reviews were published were retrieved by searching 2014 Journal Citation Reports (Thomson Reuters, 2016) in Web of Science.

The methodological quality of a systematic review reflects risk of bias or validity in its process and results. Previous studies found that the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) tool is reliable and valid [19–21]. Therefore, we used AMSTAR tool to assess the methodological quality of the included systematic reviews. The reporting quality of a systematic review refers to the clarity and transparency of its reporting, and poor reporting reduces the value and usefulness of systematic reviews [22]. We used the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist to assess the reporting quality of the included systematic reviews [23]. The methodological quality assessment using the AMSTAR tool was conducted by two independent reviewers (J.H.T. and F.S.), and the reporting quality assessment using the PRISMA tool was conducted by one reviewer (J.Z. or L.G.) and checked by a second reviewer (J.H.T. or F.S.). Any disagreements between reviewers were resolved by discussion or the involvement of a third reviewer.

To examine the agreement between the two independent reviewers in the assessment of methodological quality of systematic reviews, we calculated the agreement proportion and Cohen's kappa value for each of the 11 AMSTAR items.

### 2.3. Data analysis methods

The main characteristics and quality of systematic reviews from China vs. those from the USA were tabulated. We compared the quality of systematic reviews between China and the USA by individual items of the AMSTAR and PRISMA instrument. We calculated a summary AMSTAR score for each systematic review according to the method used by Shea et al. [20]: For each of the 11 items of the AMSTAR checklist, it was scored "1" if the answer was "Yes," and "0" if the answer was "No" or "Can't tell." Some items may be relevant mainly to meta-analysis, such as the use of funnel plot and related statistical test for assessing publication bias. It was scored "1" if a narrative discussion of risk of publication bias was available in systematic reviews when the use of funnel plot was impossible or inappropriate. The summary AMSTAR score for a systematic review was calculated by counting the number of "Yes" answers, with a possible maximum score of 11. For the assessment of reporting quality of systematic reviews, each of the 27 PRISMA items was scored "1" for full compliance, "0.5" for partial compliance, and "0" for noncompliance [15]. The summary PRISMA score for a systematic review was calculated by adding up scores assigned to each item, with a maximum score of 27.

Chi-squared test was used for differences in proportions (or Fisher's exact test if a contingency table contained a cell with five or fewer events). Two sample Wilcoxon rank-sum

test or Student *t*-test was used for differences in continuously distributed variables. The summary AMSTAR scores and the summary PRISMA scores were ranked into three groups: low (up to the 25th percentile), moderate (the interquartile range), and high (the 75th percentile and above). The association between the summary quality scores and country was calculated in either bivariate or multiple variable linear regression analyses after adjusting for factors with imbalanced distribution between China and the USA. Analyses were conducted by using data from all included systematic reviews and using data from only non-Cochrane systematic reviews. Statistical significance was defined as two sided  $P < 0.05$ . We used Stata/IC version 14.1 for statistical analyses.

## 3. Results

The search of PubMed on January 8th 2016 identified 1,073 records of systematic reviews published in 2014 from China and 796 from the USA. The references and main characteristics of the randomly selected systematic reviews are available in [Supplementary File 2](#) at [www.jclinepi.com](http://www.jclinepi.com).

### 3.1. Characteristics of included systematic reviews

The main characteristics of the included systematic reviews are summarized in [Table 1](#). Compared with systematic reviews from the USA, those from China were more likely to contain a quantitative meta-analysis, less likely to be a Cochrane systematic review, and tended to be published in journals with lower impact factors. Systematic reviews from China were more likely to have four or more coauthors, but much less likely to include coauthors from other countries. The proportion of systematic reviews with a registered review protocol was lower for systematic reviews from China, although the difference was no longer significant after excluding Cochrane systematic reviews.

There were statistically significant differences in disease conditions investigated. Compared with reviews from the USA, reviews from China were more likely to investigate cancer or tumor diseases and less likely to study mental or behavioral disorders. In terms of interventions evaluated, systematic reviews from China focused less on pharmacological interventions and more on surgical interventions and alternative medicine. Systematic reviews from China provided somewhat more significant or positive conclusions and less uncertain conclusions, although the overall difference was statistically nonsignificant.

### 3.2. Results of methodological quality

Two reviewers independently assessed the methodological quality and proportions of agreement on the initial AMSTAR assessment were greater than 65% for 10 of the 11 AMSTAR items (see [Supplementary File 3](#) at [www.jclinepi.com](http://www.jclinepi.com)).

**Table 1.** The main characteristics of the included systematic reviews

Characteristics	All systematic reviews			Non-Cochrane reviews		
	China (n = 100)	USA (n = 100)	P value	China (n = 92)	USA (n = 74)	P value
Meta-analysis (%)	97.0	77.0	***	100	74.3	***
Cochrane systematic reviews (%)	8.0	26.0	***			
Journal impact factors: median (IQR)	2.7 (1.5, 3.2)	3.7 (2.4, 6.0)	***	2.4 (1.4, 3.2)	3.0 (2.2, 4.0)	***
No. of authors (%)			***			***
1 to 3 authors	8.0	34.0		7.6	31.1	
4 to 6 authors	60.0	43.0		62.0	41.9	
7 or more authors	32.0	23.0		30.4	27.0	
Authors from 2 or more countries (%)	12.0	98.0	***	10.9	100	***
Review protocol registered (%)	12.0	29.0	**	4.4	5.4	N.S.
Type of primary outcomes (%)						
Objectively measured	42.0	42.0	N.S.	39.1	28.4	N.S.
Subjectively measured	12.0	11.0		12.0	12.2	
Not defined or unclear	46.0	47.0		48.9	59.5	
No. of RCTs included: median (IQR)	8 (6, 13)	9 (5, 16)	N.S.	8 (6, 13)	10 (6, 17)	N.S.
No. of patients included: median (IQR)	976 (514, 2124)	1552 (506, 6477)	N.S.	976 (519, 2184)	2303 (585, 12129)	**
Conditions investigated <sup>a</sup> (%)						
Circulatory	21.0	18.0	*	20.7	17.6	*
Neoplasms	24.0	8.0		26.1	9.5	
Digestive	13.0	9.0		10.9	8.1	
Endocrine/nutritional/metabolic	7.0	7.0		7.6	8.1	
Respiratory system	7.0	8.0		6.5	5.4	
Mental/behavioural	4.0	11.0		4.4	14.9	
Musculoskeletal/connective	6.0	8.0		6.5	8.1	
Genitourinary	7.0	5.0		6.5	5.4	
Any other	11.0	27.0		10.9	23.0	
Type of interventions (%)						
Pharmacological	53.0	69.0	N.S.	51.1	66.2	*
Surgical	13.0	5.0		14.1	4.1	
Device	9.0	6.0		7.6	8.1	
Disease management	6.0	6.0		6.5	6.8	
Psycho-educational	0.0	4.0		0.0	5.4	
Alternative/traditional	10.0	6.0		10.9	6.8	
Any other	9.0	4.0		9.9	2.7	
Review authors' conclusion <sup>b</sup> (%)						
Significant/positive	78.0	63.0	N.S.	80.4	68.9	N.S.
Non-significant	10.0	14.0		9.8	13.5	
Unclear/lack of evidence	12.0	23.0		9.8	17.6	

Abbreviations: ICD, International Classification of Diseases; IQR, interquartile range.

<sup>a</sup> Conditions were grouped according to ICD-10.

<sup>b</sup> Significant/positive conclusions – significant differences found or at least one intervention recommended; Non-significant conclusions – no significant differences found or no interventions recommended; Unclear conclusions – lack of evidence or neither positive nor negative, or not able to judge. Chi-square test was used for categorical variables and two sample Wilcoxon rank sum test for continuously distributed variables. N.S. =not statistically significant, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

[jclinepi.com](http://jclinepi.com)). The low agreement (28.5%) on the score for conflict of interests was mainly due to different understanding of definition for this item. All disagreements were resolved by discussion, and results of the assessment of the methodological quality are shown in [Table 2](#).

Differences between systematic reviews from China and the USA were statistically significant for seven of the 11 AMSTAR items. The methodological quality of systematic reviews from one country was not consistently lower or higher than another country for all AMSTAR items. Compared with systematic reviews from the USA, the

methodological quality of reviews from China was relatively poor in terms of a priori design, comprehensive literature search, listing both included and excluded studies, and stating sources of support in both the review and included primary studies. On the contrary, systematic reviews from China showed better quality in terms of duplicate study selection, duplicate data extraction, the assessment of scientific quality, and using scientific quality in formulating conclusions. After excluding Cochrane systematic reviews, differences in the methodological quality remained unchanged in general ([Table 2](#)).

**Table 2.** The methodological quality of systematic reviews by country – results of the AMSTAR checklist assessment

AMSTAR item	All systematic reviews			Non-Cochrane review		
	China (n = 100) %	USA (n = 100) %	P value	China (n = 92) %	USA (n = 74) %	P value
1. Was a priori design provided?	11.0	34.0	***	4.4	12.2	N.S.
2. Was there duplicate study selection and data extraction?	97.0	73.0	***	96.7	67.6	***
3. Was a comprehensive literature search performed?	78.0	91.0	*	76.1	87.8	N.S.
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	27.0	31.0	N.S.	21.7	18.9	N.S.
5. Was a list of studies (included and excluded) provided?	12.0	40.0	***	5.4	21.6	**
6. Were the characteristics of the included studies provided?	100	99.0	N.S.	100	98.7	N.S.
7. Was the scientific quality of the included studies assessed and documented?	97.0	75.0	***	97.8	66.2	***
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	80.0	51.0	***	79.4	37.8	***
9. Were the methods used to combine the findings of studies appropriate?	100	95.0	N.S.	100	93.2	*
10. Was the likelihood of publication bias assessed?	63.0	51.0	N.S.	63.0	47.3	*
11. Was the conflict of interest included?	9.0	21.0	*	5.4	14.9	*

‘Yes’ answer was used to replace ‘Not applicable’ for item 9 in systematic reviews without a meta-analysis. N.S. = not statistically significant, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

The difference in the mean summary AMSTAR score between China and the USA was statistically nonsignificant using data from all included reviews but statistically significant after excluding Cochrane reviews (Table 3). The included reviews were ranked into three groups according to summary AMSTAR scores (low, moderate, or high), and the proportions of systematic reviews belonging to each of the groups are shown in Fig. 1. The difference in proportions between the two countries was statistically significant. Compared with systematic reviews from the USA, those from China were less likely to have a low summary AMSTAR score and more likely to have a moderate summary AMSTAR score. Reviews from China were less likely to have a high AMSTAR score by using data from all systematic reviews, although it was no longer the case after excluding Cochrane reviews (Fig. 1).

### 3.3. Results of reporting quality

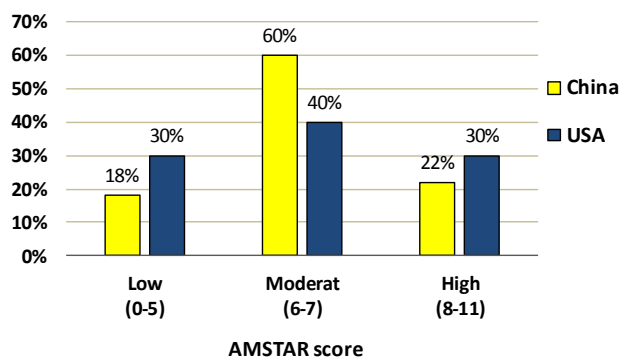
Table 4 shows proportions of systematic reviews with total compliance for each of the 27 PRISMA items. Considering all the included systematic reviews, the differences between China and the USA were statistically significant for 11 of the 27 PRISMA items. Compared with systematic reviews from the USA, those from China had a lower total compliance in reporting of review protocols, study selection methods, additional analysis methods, and funding sources. On the contrary, systematic reviews from China had a higher proportion of total compliance in reporting of titles, eligibility criteria, methods for assessing risk of bias within studies, results of risk of bias within studies, results of individual studies, results of evidence synthesis, and discussion of conclusions. After excluding Cochrane reviews, differences in reporting of titles and protocol

**Table 3.** The summary PRISMA and AMSTAR scores by country

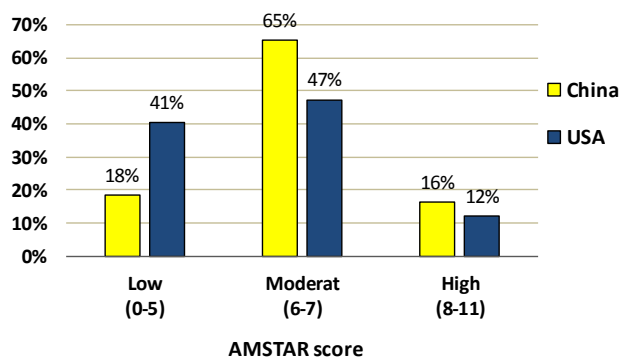
	China		The USA		Statistical significance
	n	Mean (95% CI)	n	Mean (95% CI)	
Summary AMSTAR score					
All systematic reviews	100	6.7 (6.5 to 7.0)	100	6.6 (6.1 to 7.1)	N.S.
Non-Cochrane reviews	92	6.5 (6.2 to 6.7)	74	5.7 (5.2 to 6.1)	***
Summary PRISMA score					
All systematic reviews	100	21.2 (20.7 to 21.6)	100	20.6 (19.9 to 21.3)	N.S.
Non-Cochrane reviews	92	21.0 (20.5 to 21.5)	74	19.8 (19.0 to 20.6)	*

Student t test was used for the difference in the mean summary score between China and the USA. N.S. = statistically non-significant, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

## All included systematic reviews



## Non-Cochrane systematic reviews



**Fig. 1.** The summary AMSTAR score by country. The difference in proportions between China and the USA was statistically significant ( $P = 0.016$  for all systematic reviews and  $P = 0.007$  for non-Cochrane systematic reviews).

registration between countries were no longer statistically significant, whereas differences in structured abstract and reporting of risk of bias across studies became statistically significant (Table 4).

The difference in the mean summary PRISMA score was nonsignificant using data from all included reviews and was statistically significant after excluding Cochrane reviews (Table 3). Fig. 2 shows the proportions of systematic reviews stratified into three groups according to summary PRISMA scores (low, moderate, or high). Using data from all included systematic reviews, the difference in the proportion between the two countries was statistically nonsignificant. After excluding Cochrane reviews, the overall difference was statistically significant, indicating that systematic reviews from China were less likely to have a low PRISMA score and more likely to have a high PRISMA score, compared with those from the USA (Fig. 2).

### 3.4. Results of regression analyses

Using data from all included systematic reviews, the summary AMSTAR scores were not statistically significantly associated with country in either bivariate or multiple variable linear regression analyses after adjusting for factors

with imbalanced distribution between China and the USA (Supplementary File 4 at [www.jclinepi.com](http://www.jclinepi.com)). After excluding Cochrane reviews, the summary AMSTAR scores were statistically significantly associated with country in bivariate analysis, although it was no longer significant after adjusting for other factors. Similarly, the summary PRISMA scores were not significantly associated with country when all systematic reviews were included in regression analyses. For non-Cochrane reviews, the association between the summary PRISMA score and country was statistically significant in bivariate analysis, and the association became statistically nonsignificant in multiple variable analysis (Supplementary File 4 at [www.jclinepi.com](http://www.jclinepi.com)).

## 4. Discussion

There were significant differences in characteristics of systematic reviews between the two countries, regarding the use of meta-analysis, being a Cochrane review, impact factor of journals in which they were published, and coauthors from multiple countries. The overall differences in the methodological and reporting quality of systematic reviews between China and the USA were not statistically significant after adjusting for multiple review characteristics.

Of the included systematic reviews, 8 from China and 26 from the USA were Cochrane systematic reviews. The quality of Cochrane systematic reviews was better than that of non-Cochrane reviews in the current study, which is consistent with findings from the previous studies [13,24]. Cochrane systematic reviews do not use “systematic review” or “meta-analysis” terms in titles, and all are required to register their protocols. After excluding Cochrane reviews, differences between the two countries in the reporting of titles and predefined protocols were no longer statistically significant.

Using the AMSTAR and PRISMA criteria as the “gold standard,” systematic reviews from either China or the USA need to be further improved, as systematic reviews from any other countries. For example, systematic reviews from different countries published in 2014 often failed to provide important aspects of review methods, did not search for unpublished studies, and used inappropriate statistical methods [13]. The current study found that systematic reviews from China had poor methodological quality in terms of a priori design, listing of excluded studies, and stating sources of support in both the review and included primary studies. Only 5.4% of non-Cochrane systematic reviews from China (vs. 14.9% from the USA) adequately assessed the conflict of interests in primary studies included.

Compared with systematic reviews from the USA, those from China had relatively better methodological quality in duplicate study selection and in duplicate data extraction. The number of authors of systematic reviews from China was on average larger than that from the USA, and the

**Table 4.** The reporting quality of systematic reviews by country – results of the PRISMA checklist assessment

Section	Topic	All systematic reviews			Non-Cochrane reviews			
		China (n = 100) %	USA (n = 100) %	P value	China (n = 92) %	USA (n = 74) %	P value	
Title	1. Title	88.0	72.0	**	91.3	93.2	N.S.	
Abstract	2. Structured summary	95.0	87.0	N.S.	94.6	82.4	*	
Introduction	3. Rationale	100	100		100	100		
	4. Objectives	94.0	95.0	N.S.	93.5	93.2	N.S.	
Methods	5. Protocol and registration	12.0	27.0	*	5.4	5.4	N.S.	
	6. Eligibility criteria	93.0	82.0	*	92.4	77.0	*	
	7. Information sources	100	99.0	N.S.	92.0	73.0	N.S.	
	8. Search	25.0	39.0	N.S.	19.6	20.6	N.S.	
	9. Study selection	32.0	52.0	**	27.2	46.0	*	
	10. Data collection process	71.0	69.0	N.S.	68.5	62.2	N.S.	
	11. Data items	71.0	58.0	N.S.	72.8	59.5	N.S.	
	12. Risk of bias in individual studies	95.0	68.0	***	94.6	58.1	***	
	13. Summary measures	97.0	93.0	N.S.	96.7	90.5	N.S.	
	14. Synthesis of results	98.0	93.0	N.S.	97.8	94.6	N.S.	
	15. Risk of bias across studies	68.0	68.0	N.S.	66.3	67.6	N.S.	
	16. Additional analyses	57.0	82.0	***	54.4	81.1	***	
	Results	17. Study selection	89.0	92.0	N.S.	88.0	89.2	N.S.
		18. Study characteristics	97.0	95.0	N.S.	96.7	94.6	N.S.
		19. Risk of bias within studies	64.0	51.0	***	64.0	35.1	***
20. Results of individual studies		93.0	82.0	*	92.4	77.0	**	
21. Synthesis results		98.0	87.0	***	97.8	85.1	***	
22. Risk of bias across studies		53.0	41.0	N.S.	56.5	40.5	*	
23. Additional analysis		46.0	48.0	N.S.	47.8	51.4	N.S.	
Discussion	24. Summary of evidence	100	96.0	N.S.	100	96.0	N.S.	
	25. Limitations	85.0	79.0	N.S.	83.7	77.0	N.S.	
	26. Conclusions	78.0	62.0	*	77.2	50.0	***	
Funding	27. Funding	20.0	60.0	***	18.5	58.1	***	

Proportions of total compliance for each item. *P* values were results of chi-squared testing of total compliance, partial compliance or no compliance. N.S. =not statistically significant, \* *P*<0.05, \*\* *P*<0.01, \*\*\* *P*<0.001.

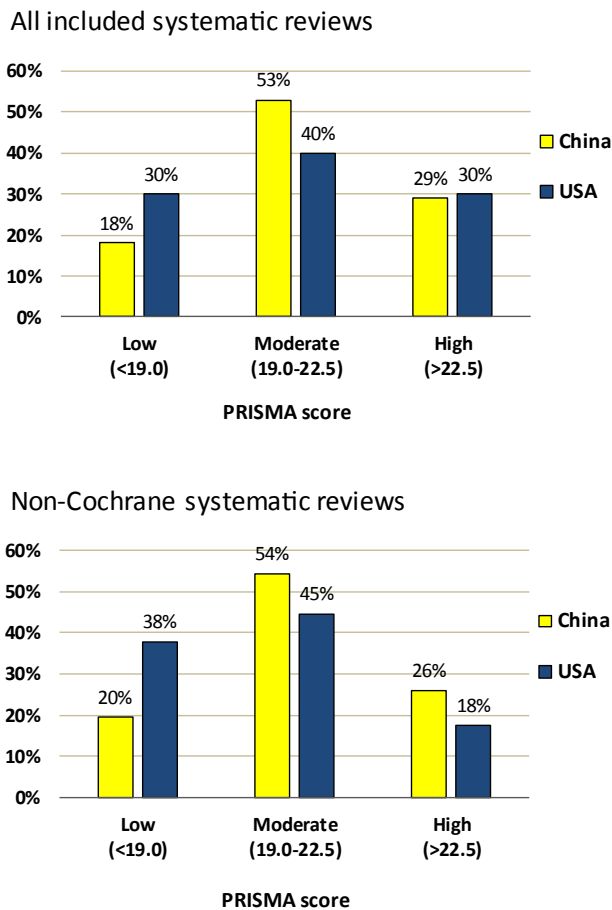
sufficient manpower is necessary to carry out duplicate study selection and data extraction. In addition, systematic reviews from China had better quality in terms of the assessment of scientific quality of studies, use of quality assessment in formulating conclusions, and assessment of publication bias. However, there are still considerable rooms for further improvement by authors from China in these items. For example, 20.6% of the included systematic reviews from China did not appropriately incorporate the scientific quality of the included studies in formulating conclusions, and 37.0% did not assess the risk of publication bias. Even there were no significant differences between the two countries, further improvement is also required. For example, literature search was not sufficiently comprehensive in 23.9%, and the status of publication (such as gray literature) was not used as an inclusion criterion in as high as 79.3% of systematic reviews from China.

Therefore, appropriate training should be provided to authors of systematic reviews in China to avoid or reduce the methodological shortcomings identified in this study. It should be emphasized that the improvement in methodological quality is also relevant to authors from the USA and likely to be relevant to systematic reviews by authors from any countries [13].

The validity and quality of findings from primary research conducted in China have been assessed in some

previous studies. For example, controlled trials of acupuncture in China reported more positive results than those from England, possibly due to publication bias [25]. Another study found that the reporting quality and validity of RCTs in China was low, compared with “gold standard” trials reported in European and North American journals [26]. More recently, Yao et al. [18] reported that quality of evidence included in meta-analyses published in Chinese language was lower than that in Cochrane systematic reviews. However, it is important to distinguish the conceptual difference between the quality of primary research and the quality of systematic reviews. Irrespective of quality of primary research studies, high-quality systematic reviews can be conducted to correctly indicate the credibility of the available evidence.

Primary research in China, as in other low- and middle-income countries (LMICs), has been rather limited in quantity and quality. For example, 78% of RCTs of interventions for major NCDs recruited patients in high-income countries, and risk of bias was higher in RCTs from LMICs [27]. Clinical and public health practice in China (as in other LMICs) will currently have to be based on research evidence mainly from high-income countries. Evidence-based health policy and clinical guidelines in China need sufficient capacity of systematic reviewing to borrow research evidence from other countries. In addition, the improved capacity in conducting



**Fig. 2.** The summary PRISMA score by country. The difference between China and the USA was statistically nonsignificant for all systematic reviews ( $P = 0.089$ ) and statistically significant for non-Cochrane reviews ( $P = 0.029$ ).

systematic reviews may also facilitate more relevant and valid primary research in China [28]. Therefore, we should celebrate the success of International Cochrane Collaboration and other evidence-based medicine efforts to increase the number and to improve the quality of systematic reviews globally during the past two decades. The concern recently raised about the redundant publication of systematic reviews [11,29,30] should be resolved by rigorous peer reviewing and editorial process [8,31].

#### 4.1. Strengths and limitations

According to our knowledge, this is the first study to compare the methodological and reporting quality of systematic reviews from China and the USA. Recent systematic reviews of RCTs on health care interventions from the two countries were randomly selected without restriction about medical field or type of interventions, so that the results would be widely generalizable and reflecting the present circumstances. Consequently, the included systematic reviews were diverse in terms of disease conditions, interventions evaluated, and other review characteristics.

Studies in future may consider to compare the quality of reviews from different countries on the same topic in terms of patients and interventions evaluated. We used regression analyses to adjust for multiple review characteristics in the comparison of the quality of systematic reviews between the two countries. The results of multiple variable analyses should be interpreted with caution because of the possible multicollinearity between independent variables.

Only the assessment of methodological quality using the AMSTAR checklist was conducted by two independent reviewers in this study, and the assessment of PRISMA reporting quality was conducted by one reviewer and checked by a second reviewer. In addition, the assessment of methodological quality of systematic reviews was based on what was reported by authors, and the actual conduct might be different. We reported results of the methodological and reporting quality of systematic reviews by checklist items and as the summary quality scores. Although the use of the summary AMSTAR score for assessing the methodological quality of systematic reviews was validated in the previous studies [20], the PRISMA checklist was not originally designed as a scored instrument [23], and further studies are required to assess the validity of the summary PRISMA score for the reporting quality of published systematic reviews. As in a previous study [20], we calculated and presented the mean AMSTAR and PRISMA summary scores in the current study. However, further studies are required to explore the appropriate statistical methods for estimating an average value of the quality scores of multiple systematic reviews.

The current study included only systematic reviews from China and the USA and assessed only systematic reviews of RCTs and published in English. Further studies are required to compare the quality of systematic reviews between other countries, published in different languages, and included observational studies. Another limitation of the current study is that the representativeness of the randomly selected systematic reviews was not assessed. The number of the included systematic reviews was based on the available time and other resources, and sample size required was not formally calculated because of no information on the meaningful difference in reporting or methodological quality of systematic reviews between countries. It may be interesting to note that the current study included a total of 100 systematic reviews from each of the two countries, twice more than the number of meta-analyses ( $n = 50$ ) from each of the two countries included in a previous study [12].

## 5. Conclusions

The overall methodological and reporting quality of systematic reviews by authors from China were similar to those from the USA, although the quality of systematic reviews from both countries could be further improved. Identified shortcomings in methodological and reporting quality of published systematic



reviews should be taken into consideration in further training of authors of systematic reviews in the relevant countries.

## Acknowledgments

Authors' contributions: K.H.Y. and F.S. conceived initial concept. F.S. and J.H.T. developed review protocol. F.S. conducted literature and study selection. J.H.T., J.Z., L.G., and F.S. extracted and checked data from the included systematic reviews and assessed reporting and methodological quality of systematic reviews. F.S. performed data analysis and drafted the manuscript. All authors commented on the manuscript.

## Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2016.12.004>.

## References

- [1] Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126: 376–80.
- [2] Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, 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;312:171–9.
- [3] Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof* 2002;25(1):12–37.
- [4] Summerskill W. Cochrane collaboration and the evolution of evidence. *Lancet* 2005;366:1760.
- [5] Zhang M, Li Y. Cochrane collaboration in China 1996–2013. *J Evid Based Med* 2014;7:22–5.
- [6] Mallett S, Clarke M. How many Cochrane reviews are needed to cover existing evidence on the effects of health care interventions? *ACP J Club* 2003;139(1):A11.
- [7] Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7(9):e1000326.
- [8] Fuhr U, Hellmich M. Channeling the flood of meta-analyses. *Eur J Clin Pharmacol* 2015;71:645–7.
- [9] Tebala GD. What is the future of biomedical research? *Med Hypotheses* 2015;85(4):488–90.
- [10] Ioannidis JP. Can too many systematic reviews and meta-analyses do harm? Plenary II: Information overload: are we part of the problem or part of the solution?. The 24th Cochrane Colloquium. Vienna, 2015.
- [11] Riaz IB, Khan MS, Riaz H, Goldberg RJ. Disorganized systematic reviews and meta-analyses: time to systematize the conduct and publication of these study overviews? *Am J Med* 2016;129:339.e11–8.
- [12] Ioannidis JP, Chang CQ, Lam TK, Schully SD, Khoury MJ. The geometric increase in meta-analyses from China in the genomic era. *PLoS One* 2013;8(6):e65602.
- [13] Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13(5): e1002028.
- [14] Jin YH, Ma ET, Gao WJ, Hua W, Dou HY. Reporting and methodological quality of systematic reviews or meta-analyses in nursing field in China. *Int J Nurs Pract* 2014;20:70–8.
- [15] Li JL, Ge L, Ma JC, Zeng QL, Yao L, An N, et al. Quality of reporting of systematic reviews published in “evidence-based” Chinese journals. *Syst Rev* 2014;3:58.
- [16] Ma B, Guo J, Qi G, Li H, Peng J, Zhang Y, et al. Epidemiology, quality and reporting characteristics of systematic reviews of traditional Chinese medicine interventions published in Chinese journals. *PLoS One* 2011;6(5):e20185.
- [17] Ma B, Qi GQ, Lin XT, Wang T, Chen ZM, Yang KH. Epidemiology, quality, and reporting characteristics of systematic reviews of acupuncture interventions published in Chinese journals. *J Altern Complement Med* 2012;18:813–7.
- [18] Yao L, Sun R, Chen YL, Wang Q, Wei D, Wang X, et al. The quality of evidence in Chinese meta-analyses needs to be improved. *J Clin Epidemiol* 2016;74:73–9.
- [19] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- [20] 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. *J Clin Epidemiol* 2009;62:1013–20.
- [21] Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol* 2015;68: 574–83.
- [22] Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4(3):e78.
- [23] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009; 62:e1–34.
- [24] Petticrew M, Wilson P, Wright K, Song F. Quality of Cochrane reviews. Quality of Cochrane reviews is better than that of non-Cochrane reviews. *BMJ* 2002;324:545.
- [25] 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;19:159–66.
- [26] Zhang D, Freemantle N, Cheng KK. Are randomized trials conducted in China or India biased? A comparative empirical analysis. *J Clin Epidemiol* 2011;64:90–5.
- [27] Fan H, Song F. An assessment of randomized controlled trials (RCTs) for non-communicable diseases (NCDs): more and higher quality research is required in less developed countries. *Sci Rep* 2015;5: 13221.
- [28] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156–65.
- [29] Bolland MJ, Grey A. A case study of discordant overlapping meta-analyses: vitamin d supplements and fracture. *PLoS One* 2014; 9(12):e115934.
- [30] Helfer B, Prosser A, Samara MT, Geddes JR, Cipriani A, Davis JM, et al. Recent meta-analyses neglect previous systematic reviews and meta-analyses about the same topic: a systematic examination. *BMC Med* 2015;13:82.
- [31] D'Agostino RB Jr. Editorial commentary: meta-analyses are on the rise, but are we learning more? *Arthroscopy* 2016;32:538–9.