BMJ Open

Epidemiology, quality and reporting characteristics of meta-analyses of observational studies published in Chinese journals

Zhe-wen Zhang,[1] Juan Cheng,[2] Zhuan Liu,[1] Ji-chun Ma,[3] Jin-long Li,[3] Jing Wang,[3] Ke-hu Yang[4]

For numbered affiliations see end of article.

Correspondence to
Ke-hu Yang;
yebm123@163.com

ABSTRACT

Objective: The aim of this study was to examine the epidemiological and reporting characteristics as well as the methodological quality of meta-analyses (MAs) of observational studies published in Chinese journals.

Methods: 5 Chinese databases were searched for MAs of observational studies published from January 1978 to May 2014. Data were extracted into Excel spreadsheets, and Meta-analysis of Observational Studies in Epidemiology (MOOSE) and Assessment of Multiple Systematic Reviews (AMSTAR) checklists were used to assess reporting characteristics and methodological quality, respectively.

Results: A total of 607 MAs were included. Only 52.2% of the MAs assessed the quality of the included primary studies, and the retrieval information was not comprehensive in more than half (85.8%) of the MAs. In addition, 50 (8.2%) MAs did not search any Chinese databases, while 126 (20.8%) studies did not search any English databases. Approximately 41.2% of the MAs did not describe the statistical methods in sufficient details, and most (95.5%) MAs did not report on conflicts of interest. However, compared with the before publication of the MOOSE Checklist, the quality of reporting improved significantly for 20 subitems after publication of the MOOSE Checklist, and 7 items of the included MAs demonstrated significant improvement after publication of the AMSTAR Checklist (p<0.05).

Conclusions: Although many MAs of observational studies have been published in Chinese journals, the reporting quality is questionable. Thus, there is an urgent need to increase the use of reporting guidelines and methodological tools in China; we recommend that Chinese journals adopt the MOOSE and AMSTAR criteria.

Strengths and limitations of this study

- Our study was the first to examine the compliance of Chinese observational study meta-analyses using the Meta-analysis of Observational Studies in Epidemiology reporting guidelines and the Assessment of Multiple Systematic Reviews tool for assessing methodological quality.
- This study included a comprehensive literature search using five Chinese databases to ensure a high degree of representativeness.
- In addition, this study included only meta-analyses published in Chinese journals, whereas Chinese investigators increasingly publish articles in international journals.

controlled trials (RCTs) would provide stronger evidence than those not including RCTs.[2 3] However, in many situations, randomised controlled designs are not feasible and only data from observational studies are available. Therefore, observational studies have an important role in answering questions related to treatment effectiveness and disease aetiology.

Owing to the lack of randomisation, observational studies are inherently more prone to potential biases.[4 5] For instance, case–control studies are always retrospective in nature, which increases the potential for incomplete and biased data collection. Therefore, it is more important to describe exactly the methodology that led to the generation of results from meta-analyses of observational studies.

The Meta-Analysis of Observational Studies in Epidemiology (MOOSE) checklist and the Assessment of Multiple Systematic Reviews (AMSTAR) tool were first introduced and published in China in 2010.[6 7] Over the past decades, many studies have described the quality and reporting characteristics in

INTRODUCTION

Meta-analysis as a statistical and scientific tool has grown immensely popular over the past decade.[1] Several studies have considered that meta-analyses including only randomised

multidisciplinary clinical research topics, but these studies did not include information about epidemiological characteristics or methodological quality based on the meta-analyses of observational studies in China.[8–10] The aim of this study is to describe the epidemiological and reporting characteristics, as well as the methodological quality, of the meta-analyses of observational studies published in Chinese journals, using the most up-to-date assessment tools.

METHODS
Data sources and searches
Five Chinese databases (Chinese Biomedical Literature database (CBM), Chinese Science Citation Database (CSCD), VIP information (Chinese Scientific Journals database), China National Knowledge Infrastructure (CNKI) and WANFANG database (Chinese Medicine Premier)), were searched from inception through May 2014 (see online supplementary file 1). The search terms included 'review', 'meta-analysis', 'systematic review', 'pooled analysis', 'overview', 'cohort', 'case control' and 'cross sectional'. The search was limited to the following criteria: MAs of the article type and one of three main study designs, including cohort, case–control and cross-sectional. The search was limited to human studies. Editorials, letters, conferences and meeting abstracts were excluded. Then, the full texts of the potentially eligible studies were retrieved and further evaluated. The references of retrieved articles were also searched.

Data collection and analysis
Study reports were grouped according to the year that the two checklists were introduced in China: 2009 and earlier (prepublication) or 2010–2014 (postpublication). Articles were scored as 'yes' if they were reported in enough detail to allow the reader to judge that the definition had been met. An article was scored as 'partially/cannot tell' only when the report was incomplete or unclear. Articles were coded as 'no' when the checklist item was not reported. We also collected information regarding the risk of bias tools and methods used to search Chinese journals.

To enhance the reviewers' inter-rater agreement, we evaluated 20 papers (not included in the study sample) in a pilot test of the database prior to starting the data abstraction process. Proper scoring of each item in the database was discussed in detail. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guideline was also followed (see online supplementary file 2). Z-wZ and JC searched the literature. ZL, J-cM JC, J-lL and JW participated in data extraction and quality assessment of the MAs, with guidance from K-hY. Intraclass correlation coefficients (ICCs) were used to assess inter-rater reliability within each item.[11] The χ^2 test was used to compare the quality of MAs published in journals cited by CSCD and

non-CSCD. Statistical significance was considered when $p < 0.05$. Data analysis was performed with SPSS V.13.0.

RESULTS
Search
A total of 2930 potentially relevant reports from the databases were identified for review. The screening process excluded 1977 reviews due to duplication or the absence of MAs. Another 346 reviews that were not MAs of observational studies were excluded after examination of the full texts. Finally, a total of 607 MAs were considered to be eligible for our study (figure 1 and online supplementary file 3).

Descriptive characteristics
The first related methods of MAs concerning observational studies were published in China in 1995, and the overall number of published MAs has subsequently increased. The 607 MAs included were published in 265 different Chinese journals. Less than one-third (28.5%) of the MAs were supported by the government. The most common conditions studied included diseases of neoplasms (43%) and the circulatory system (17%). The number of authors ranged from 1 to 11 with a median of four authors. Less than one-fifth (18.9%) of the MAs were cited by the CSCD. In addition, 85% of the articles included the term 'meta-analyses' in the title. None of the MAs had been updated from a previous review (table 1).

Risk of bias instruments
Only 52.2% of the MAs reported that they assessed the quality of the included primary studies. Of these, 39 (6.3%) MAs used the Newcastle-Ottawa scale (NOS); 19 (3.1%) MAs used the critical appraisal skill programme (CASP); 12 (2%) MAs used the Cochrane Collaboration scale (CC); 29 (4.8%) MAs used the Strengthening the Reporting of Observational Studies in Epidemiology
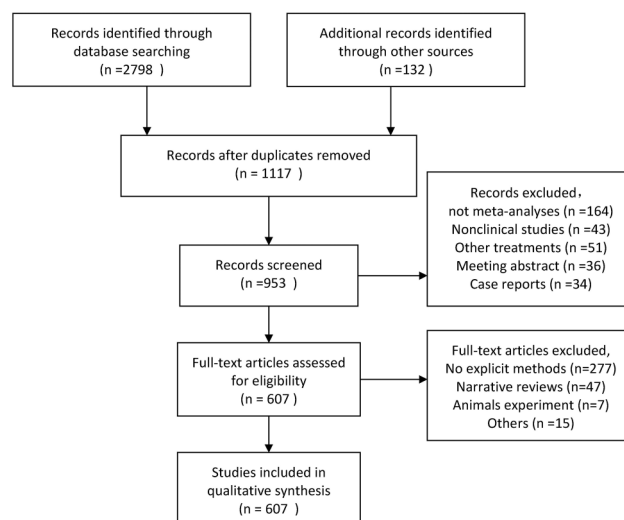

Figure 1 Flow chart of systematic search.

Table 1 Descriptive characteristics of included MAs

Category	All meta-analyses n=607
Total number of journals	265
Funding source (yes)	
Government	173 (28.5)
Industrial	0 (0)
Other	39 (6.3)
Common ICD-10	
Neoplasms	261 (43)
Diseases of the digestive system	26 (4.3)
Certain infectious and parasitic diseases	28 (4.6)
Disease of the circulatory system	103 (17)
Diseases of the musculoskeletal system and connective tissue	14 (2.3)
Diseases of the blood and blood forming organs and immune mechanism	12 (1.9)
Endocrine, nutritional and metabolic diseases	34 (5.6)
Diseases of the respiratory system	16 (2.6)
Diseases of the genitourinary system	14 (2.3)
Congenital malformations, deformations and chromosomal abnormalities	3 (0.5)
Injury, poisoning and certain other consequences of external causes	61 (10.1)
Pregnancy, childbirth and puerperium	22 (3.6)
Certain conditions originating in the perinatal period	6 (0.9)
Diseases of the eye and adnexa	7 (1.2)
Number of authors, median (IQR)	4 (1–11)
Number of included studies, median (IQR)	
Cohort, n=24	13 (2–215)
Case–controlled, n=488	12 (2–90)
Cohort + case–controlled, n=65	11 (1–189)
Cross-sectional, n=10	16 (2–63)
Number of participants in included studies, median (IQR)	
Cohort, n=89	561 279 (76–7 069 228)
Case–controlled, n=553	4575 (72–105 293)
Cross-sectional, n=10	78 (3–11 215)
Types of clinical study (yes)	
Therapy	31 (5.1)
Aetiology	576 (94.9)
Types of models in included studies (yes)	
Fixed model	107 (17.6)
Random model	164 (27.2)
Not reported	336 (55.2)
Indexed in CSCD, yes n (%)	18.9
Update of a previous review: yes n (%)	0 (0)

CSCD, Chinese Science Citation Database; ICD-10, International Classification of Diseases 10th edition; MA, meta-analyses.

(STROBE); and 218 (36%) MAs presented either the reference for the scale or used unnamed scales. Among the 607 included MAs, 548 (90.3%) provided the name and version of the statistical software employed, including particulars of any special features used (table 2).

Searching details for studies

Within the included studies, the median number of databases used was 4, with a range of 0 to 16. Regarding the Chinese-language databases, the most commonly searched database was CNKI (65.7%), followed by CBM (51.7%) and VIP (49.4%). PUBMED was the most commonly searched English-language database (66%); the second most common database was EMBASE (28.7%), followed by the Cochrane Library (16.8%). A total of 50

Table 2 Tools of quality assessment in included meta-analyses

Quality assessment	Number (%), of n=607	95% CI
NOS	39 (6.3)	4.7 to 8.7
CASP	19 (3.1)	2 to 4.9
Cochrane scale	12 (2.0)	1.1 to 3.4
STROBE*	29 (4.8)	3.3 to 6.8
Others	218 (36.0)	32.2 to 39.8
Not reported/not performed	290 (47.8)	43.8 to 51.8

*The STROBE was not a tool for assessing the quality of published observational studies.
CASP, critical appraisal skill programme; NOS, Newcastle-Ottawa scale; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

(8.2%) MAs did not search any Chinese database, and 126 (20.8%) MAs did not search any English language database. In addition, 559 (92.1%) revealed the search terms (some terms given but not all), but only 87 (14.3%) studies presented the search strategy (search terms and Boolean operators; table 3).

AMSTAR checklist (current edition) assessment

Table 4 shows the summary of results for the risk of bias of all MAs. Compliance with the AMSTAR checklist

Table 3 Search details reported by included meta-analyses

Assessment item	Number (%)
Number of databases searched, median (range)	4 (0–16)
Chinese databases searched	
China National Knowledge Infrastructure (CNKI)	464 (76.4)
Chinese Biomedical (CBM) literature database	314 (51.7)
VIP information (Chinese Scientific Journals database)	300 (49.4)
WANFANG database (Chinese Medicine Premier)	277 (45.6)
Others	34 (5.6)
English-language databases searched	
PubMed	466 (66)
EMBASE	174 (28.7)
Cochrane Library	102 (16.8)
OVID	58 (9.6)
SCI (Web OF Science)	42 (6.9)
Springerlink	39 (6.4)
Elsevier Science	37 (6.1)
Others	104 (17.1)
Number of Chinese databases	
0	50 (8.2)
1	111 (18.3)
2	156 (25.7)
3	178 (29.3)
>3	112 (18.5)
Number of English databases	
0	126 (20.8)
1	182 (30)
2	157 (25.9)
3	88 (14.5)
>3	54 (8.8)
Was the strategy given in full?	
Search strategy (search terms and Boolean operators)	87 (14.3)
Partial (eg, some terms given but not all)	559 (92.1)
Not report	15 (2.5)
Other resources searched	
Reference sections of retrieve articles	489 (80.5)
Conference abstracts/posters	336 (55.4)
Assistant retrieval methods	
Manual searching (eg, reference, conference)	411 (67.7)
Search engine (eg, Google scholar)	41 (6.8)

items ranged from 4.5 to 75.8. The overall agreement among reviewers for evaluation using the AMSTAR Checklist was moderate (ICC=0.81; 95% CI 0.71 to 0.89). Six AMSTAR items (1, 2, 3, 4, 5 and 11) were reported in less than 50% of the total reports. No significant difference was found for the MAs published in journals cited by CSCD versus non-CSCD. Compared to studies published before 2010, there was an increase in seven items (1, 2, 3, 5, 6, 7 and 10) on the AMSTAR checklist (p<0.05; table 4).

MOOSE checklist (current edition) assessment

Table 5 shows the proportion of all MAs reporting each item in the MOOSE checklist. Compliance with the MOOSE checklist items ranged from 0% to 96.7%. The overall agreement among reviewers for evaluation with the MOOSE checklist was also moderate (ICC=0.79; 95% CI 0.68 to 0.87). Fourteen MOOSE checklist subitems (2, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 20, 29 and 35) were mentioned in less than 50% of the total reports, and four of these subitems (7, 13, 14 and 16) were included in less than 10% of the reports. There was also no statistically significant difference in the source of journals cited by CSCD versus non-CSCD. In addition, the quality of reporting demonstrated significant improvement regarding the background (item 3), search strategy (items 7, 8, 9, 10, 12, 14 and 15), methods (items 17, 18, 19, 20, 21, 22 and 24), results (items 25, 27 and 28) and discussion (items 30 and 31). However, no study provided the name and version of the search software employed (subitem 11), and there was no mention of the special features used (table 5).

DISCUSSION

Our study shows that large numbers of MAs of observational studies have recently been conducted, with 607 publications identified in Chinese journals. This study was the first to examine the compliance of Chinese observational study MAs using the MOOSE reporting guidelines and the AMSTAR tool for assessing methodological quality.

This study found that the methodological quality of Chinese MAs is poor. In particular, we found the retrievals were not comprehensive and lacked bias assessments in the majority of the MAs that we examined. Reporting the details of the search strategy is a requirement for MAs, as this information facilitates an assessment of comprehensiveness and ensures reproducibility.[12] This study demonstrated that 85.8% of the MAs examined did not perform comprehensive literature searches; for example, only 14.3% of the studies presented their search strategy; 15.7% of the studies included searches of grey literature; and 67.7% of the studies used manual retrieval. Moreover, the lack of detailed retrieval strategies and qualifications of the searchers (ie, librarians and investigators) should also be noted. Ma *et al*[13] reported that 59.1% of Chinese SRs of acupuncture

Table 4 AMSTAR assessment of methodological characteristics (n=607)

Category	Yes (%) All n=607	≤2009 n=230	2010–2014 n=377	Partially/cannot tell (%) All n=607	≤2009 n=230	2010–2014 n=377	No (%) All n=607	≤2009 n=230	2010–2014 n=377	p Values
1. Was an 'a priori' design provided?	227 (37.4)	69 (30.0)	158 (41.9)	310 (51.1)	134 (58.3)	176 (46.7)	70 (11.5)	27 (11.7)	43 (11.4)	0.010
2. Was there duplicate study selection and data extraction?	255 (42.0)	65 (28.3)	190 (50.4)	13 (2.2)	6 (2.6)	7 (1.9)	339 (55.8)	159 (69.1)	180 (47.7)	0.000
3. Was a comprehensive literature search performed?	86 (14.2)	22 (9.6)	64 (16.9)	392 (64.6)	141 (61.3)	251 (66.6)	129 (21.2)	67 (29.1)	62 (16.5)	0.000
4. Was the status of publication (ie, grey literature) used as an inclusion criterion?	95 (15.7)	37 (16.1)	58 (15.4)	0 (0.0)	0 (0.0)	0 (0.0)	512 (84.3)	193 (83.9)	319 (84.6)	0.817
5. Was a list of studies (included and excluded) provided?	63 (10.4)	20 (8.7)	43 (11.4)	96 (15.8)	50 (21.7)	46 (12.2)	448 (73.8)	160 (69.6)	288 (76.4)	0.006
6. Were the characteristics of the included studies provided?	460 (75.8)	158 (68.7)	302 (80.1)	12 (1.9)	6 (2.6)	6 (1.6)	135 (22.3)	66 (28.7)	69 (18.3)	0.006
7. Was the scientific quality of the included studies assessed and documented?	317 (52.2)	96 (41.7)	221 (58.6)	0 (0.0)	0 (0.0)	0 (0.0)	290 (47.8)	134 (58.3)	156 (41.4)	0.000
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	357 (58.8)	124 (53.9)	233 (61.8)	23 (3.8)	8 (3.5)	15 (3.9)	227 (37.4)	98 (42.6)	129 (34.3)	0.117
9. Were the methods used to combine the findings of studies appropriate?	436 (71.8)	169 (73.5)	267 (70.8)	150 (24.7)	54 (23.5)	97 (25.7)	21 (3.5)	7 (3.0)	13 (3.5)	0.055
10. Was the likelihood of publication bias assessed?	374 (61.6)	110 (47.8)	264 (70.0)	11 (1.8)	6 (2.6)	5 (1.3)	222 (36.6)	114 (49.6)	108 (28.7)	0.000
11. Was the conflict of interest stated?	27 (4.5)	8 (3.5)	19 (5.0)	0 (0.0)	0 (0.0)	0 (0.0)	580 (95.5)	222 (96.5)	358 (95.0)	0.356

Table 5 MOOSE assessment of reporting characteristics (n=607)

Category	Yes (%) All n=607	≤2009 n=230	2010–2014 n=377	Partially/cannot tell (%) All n=607	≤2009 n=230	2010–2014 n=377	No (%) All n=607	≤2009 n=230	2010–2014 n=377	p Values
Reporting of background should include										
1. Problem definition	536 (88.3)	198 (86.1)	338 (89.7)	28 (4.6)	14 (6.1)	14 (3.7)	43 (7.1)	18 (7.8)	25 (6.6)	0.080
2. Hypothesis statement	279 (45.9)	94 (40.9)	185 (49.1)	165 (27.2)	64 (27.8)	101 (26.8)	163 (26.9)	72 (31.3)	91 (24.1)	0.087
3. Description of study outcome (s)	376 (61.9)	129 (56.1)	247 (65.5)	79 (13.0)	45 (19.6)	34 (9.1)	152 (25.1)	56 (24.3)	96 (25.4)	0.001
4. Type of exposure or intervention used	430 (70.9)	154 (66.9)	276 (73.2)	2 (0.3)	0 (0.0)	2 (0.5)	175 (28.8)	76 (33.1)	99 (26.3)	0.117
5. Type of study designs used	202 (33.3)	70 (30.4)	132 (35.1)	0 (0.0)	0 (0.0)	0 (0.0)	405 (66.7)	160 (69.5)	245 (64.9)	0.246
6. Study population	234 (38.6)	86 (37.4)	148 (39.3)	0 (0.0)	0 (0.0)	0 (0.0)	373 (61.4)	144 (62.6)	229 (60.7)	0.647
Reporting of search strategy should include										
7. Qualifications of searchers (ie, librarians and investigators)	12 (1.9)	1 (0.4)	11 (2.9)	0 (0.0)	0 (0.0)	0 (0.0)	595 (98.1)	229 (99.6)	366 (97.1)	0.033
8. Search strategy, including time period included in the synthesis and keywords	167 (27.5)	38 (16.5)	129 (34.2)	396 (65.2)	167 (72.6)	229 (60.7)	44 (7.3)	25 (10.9)	19 (5.1)	0.000
9. Effort to include all available studies, including contact with authors	212 (34.9)	53 (23.1)	159 (42.2)	0 (0.0)	0 (0.0)	0 (0.0)	395 (65.1)	177 (76.9)	218 (57.8)	0.000
10. Databases and registries searched	401 (66.1)	134 (58.3)	267 (70.8)	177 (29.2)	78 (33.9)	99 (26.3)	29 (4.7)	18 (7.8)	11 (2.9)	0.001
11. Search software used, name and version, including special features used (ie, explosion)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	607 (100.0)	230 (100.0)	377 (100.0)	1
12. Use of manual searching (ie, reference lists of obtained articles)	125 (20.6)	36 (15.7)	89 (23.6)	0 (0.0)	0 (0.0)	0 (0.0)	482 (79.4)	194 (84.3)	288 (76.4)	0.019
13. List of citations located and those excluded, including justification	43 (7.1)	12 (5.2)	31 (8.2)	57 (9.4)	20 (8.7)	37 (9.8)	507 (83.5)	198 (86.1)	309 (82.0)	0.060
14. Method of addressing articles published in languages other than English	21 (3.5)	2 (0.9)	19 (5.1)	0 (0.0)	0 (0.0)	0 (0.0)	586 (96.5)	228 (99.1)	358 (94.9)	0.006
15. Method of handling abstracts and unpublished studies	110 (18.1)	25 (10.9)	85 (22.5)	89 (14.7)	10 (4.3)	79 (20.9)	408 (67.2)	195 (84.8)	213 (56.4)	0.000
16. Description of any contact with authors	26 (4.3)	9 (3.9)	17 (4.5)	0 (0.0)	0 (0.0)	0 (0.0)	581 (95.7)	221 (96.1)	360 (95.5)	0.725
Reporting of methods should include										
17. Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	427 (70.3)	148 (64.3)	279 (74.1)	0 (0.0)	0 (0.0)	0 (0.0)	180 (29.7)	82 (35.7)	98 (25.9)	0.011
18. Rationale for the selection and coding of data (ie, sound clinical principles or convenience)	430 (70.8)	149 (64.8)	281 (74.5)	0 (0.0)	0 (0.0)	0 (0.0)	177 (29.2)	81 (35.2)	96 (25.5)	0.010

Table 5 Continued

Category	Yes (%) All n=607	≤2009 n=230	2010–2014 n=377	Partially/cannot tell (%) All n=607	≤2009 n=230	2010–2014 n=377	No (%) All n=607	≤2009 n=230	2010–2014 n=377	p Values
19. Documentation of how data were classified and coded (ie, multiple raters, blinding and inter-rater reliability)	405 (66.7)	136 (59.1)	269 (71.4)	0 (0.0)	0 (0.0)	0 (0.0)	202 (33.3)	94 (40.9)	108 (28.6)	0.002
20. Assessment of confounding (ie, comparability of cases and controls in studies where appropriate)	288 (47.5)	95 (41.3)	193 (51.2)	114 (18.8)	42 (18.3)	72 (19.1)	205 (33.7)	93 (40.4)	112 (29.7)	0.020
21. Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results	413 (68.1)	145 (63.1)	268 (71.1)	0 (0.0)	0 (0.0)	0 (0.0)	194 (31.9)	85 (36.9)	109 (28.9)	0.039
22. Assessment of heterogeneity	355 (58.5)	102 (44.3)	253 (67.1)	200 (32.9)	110 (47.8)	90 (23.8)	52 (8.6)	18 (7.9)	34 (9.1)	0.000
23. Description of statistical methods (ie, complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose–response models, or cumulative meta-analysis) in sufficient detail to be replicated	305 (50.2)	108 (46.9)	197 (52.3)	250 (41.2)	100 (43.5)	150 (39.8)	52 (8.6)	22 (9.6)	30 (7.9)	0.427
24. Provision of appropriate tables and graphics	537 (88.5)	195 (84.8)	342 (90.7)	0 (0.0)	0 (0.0)	0 (0.0)	70 (11.5)	35 (15.2)	35 (9.3)	0.026
Reporting of results should include										
25. Graphic summarising individual study estimates and overall estimate	310 (51.1)	78 (33.9)	232 (61.5)	107 (17.6)	51 (22.2)	56 (14.9)	190 (31.3)	101 (43.9)	89 (23.6)	0.000
26. Table giving descriptive information for each study included	305 (50.2)	110 (47.8)	195 (51.7)	140 (23.1)	57 (24.8)	83 (22.1)	63 (27.4)	99 (26.2)	112 (29.6)	0.098
27. Results of sensitivity testing (ie, subgroup analysis)	442 (72.8)	152 (66.1)	290 (76.9)	0 (0.0)	0 (0.0)	0 (0.0)	165 (27.2)	78 (33.9)	87 (23.1)	0.004
28. Indication of statistical uncertainty of findings	587 (96.7)	217 (94.3)	370 (98.1)	0 (0.0)	0 (0.0)	0 (0.0)	20 (3.3)	13 (5.7)	7 (1.9)	0.011
Reporting of discussion should include										
29. Quantitative assessment of bias (ie, publication bias)	246 (40.5)	91 (39.6)	155 (41.1)	107 (17.6)	41 (17.8)	66 (17.5)	254 (41.9)	98 (42.6)	156 (41.4)	0.930
30. Justification for exclusion (ie, exclusion of non–English-language citations)	307 (50.6)	59 (25.7)	248 (65.8)	0 (0.0)	0 (0.0)	0 (0.0)	300 (49.4)	171 (74.3)	129 (34.2)	0.000

Continued

Table 5 Continued

Category	Yes (%)			Partially/cannot tell (%)			No (%)			p Values
	All n=607	≤2009 n=230	2010–2014 n=377	All n=607	≤2009 n=230	2010–2014 n=377	All n=607	≤2009 n=230	2010–2014 n=377	
31. Assessment of quality of included studies	317 (51.3)	94 (40.9)	223 (59.2)	0 (0.0)	0 (0.0)	0 (0.0)	290 (48.7)	136 (59.1)	154 (40.8)	0.000
Reporting of conclusions should include										
32. Consideration of alternative explanations for observed results	369 (60.8)	142 (61.7)	227 (60.2)	0 (0.0)	0 (0.0)	0 (0.0)	238 (39.2)	88 (38.3)	150 (39.8)	0.709
33. Generalisation of the conclusions (ie, appropriate for the data presented and within the domain of the literature review)	411 (67.7)	152 (66.1)	259 (68.7)	0 (0.0)	0 (0.0)	0 (0.0)	196 (32.3)	78 (33.9)	118 (31.3)	0.504
34. Guidelines for future research	304 (50.1)	125 (54.3)	179 (47.5)	0 (0.0)	0 (0.0)	0 (0.0)	303 (49.9)	105 (45.7)	198 (52.5)	0.101
35. Disclosure of funding source	170 (28.0)	61 (26.5)	109 (28.9)	0 (0.0)	0 (0.0)	0 (0.0)	437 (72.0)	169 (73.5)	268 (71.1)	0.069

interventions published in Chinese journals did not perform comprehensive literature searches and that 97.7% did not include searches of grey literature or ongoing studies. Moreover, the lack of a comprehensive search was clearly the weakest item in the identified MAs in Chinese journals.

Risk of bias is important because poor methodological quality can lead to a biased estimate. In the present research, nearly one-half of the studies did not mention how the quality of included primary studies was assessed. In addition, 29 (4.8%) studies used the STROBE criteria. However, it should be noted that the STROBE criteria were not developed as a tool for assessing the quality of published observational studies; instead, the STROBE criteria were developed solely to provide guidance on how to report observational research.[14] Similarly, Bruno et al[15] reported that about half of the systematic reviews and meta-analyses used STROBE inappropriately, as a methodological quality assessment tool. In some instances, specific checklists for observational studies were used, including the NOS and CASP, which have been shown to be generally useful for assessing the quality of non-randomised studies despite some limitations.[16 17] These assessments serve to identify the strengths and limitations of included studies, including the quality of strength of the evidence for a given outcome. The NOS has been endorsed for use in systematic reviews of non-randomised studies by the Cochrane Collaboration, specifically for cohort and case–control studies. CASP is an instrument for the appraisal of systematic reviews based on 10 questions for addressing the key components of methodological quality. Therefore, to obtain valuable findings from observational study MAs, adequate assessment based on the correct study design is essential.

In addition, many studies did not report key aspects of MAs methodologies, which reduces confidence in the results and impairs the conclusion. For example, more than half of the studies reported an 'a priori' design, and another 11.5% of the studies did not reveal their design information. The most common means of assessing publication bias was by funnel plot, and more than one-third of the studies did not consider or assess publication bias despite considerable evidence for its existence and its potential influence on the MA results. Only 4.5% of the studies stated conflicts of interest; for example, Barnes and Bero[18] reported that funding sources may have influenced the outcomes and quality of the research. These important methodology components must be considered in future research.

Accurate reporting is essential to maintain a clear scientific record, which can then be used for the synthesis of existing evidence, clinical decision-making and health policy determination. Groenwold et al[19] reported that the quality of reporting on confounding in observational studies was rather poor, even in high-impact general medical journals. Our studies showed that less than 50% of the included studies assessed confounding. As it cannot be guaranteed that known and unknown

confounding factors are distributed equally among the observation groups, results of this type are susceptible to distortions. Therefore, clinicians reading the reports of MAs must be able to appraise the method and validity of the study to confidently interpret the results.

As mentioned above, we found that the quality of reporting regarding search strategies and methods significantly improved after publication of the MOOSE checklist. However, this observation is prone to many biases and could simply represent improvements in research methods over time. Nonetheless, room for improvement still exists. For example, approximately one-half of the studies did not present risk bias assessment results, which could have affected the cumulative evidence. This is despite the fact that many studies have previously shown the importance of assessing bias heterogeneity across studies.[20 21] Disappointingly, 41.2% of the studies did not describe the statistical methods in sufficient detail; in fact, some of the studies did not explore the reasons for statistical heterogeneity and simply pooled results using a random effects model to account for heterogeneity. These shortcomings may have led to incorrect or inappropriate interpretations of the results.

Panic *et al*[22] reported that the endorsement of PRISMA resulted in increase of both quality of reporting and methodological quality. Our studies showed that less than one-fifth of the included studies were indexed in the Chinese Science Citation Database (CSCD), which is similar to the Science Citation Index. The reason may be the overall poor quality of work and many deficits in reporting in the same field in the Chinese MAs. Therefore, broader promotion of methodological quality guidelines is a necessary step in enhancing dissemination and implementation of AMSTAR and MOOSE.

The strengths of this study include its comprehensive literature search using five Chinese databases, to ensure a high degree of representativeness. In addition, both the eligibility process and data extraction were conducted by two independent investigators, with a third investigator providing quality evaluation. Nonetheless, there were some limitations in our current study. First, in this study, the terms 'meta-analysis', 'systematic review' and 'pooled analysis' were used, although some potentially eligible MAs may not have included these terms in their publications. Second, this study included only MAs published in Chinese journals, whereas Chinese investigators increasingly publish articles in international journals. Third, our studies relied on reporting from authors, and it is possible that the authors may have omitted important details from their reports or that the peer-review process resulted in the removal of key information from these reviews.

CONCLUSION

The goal of the present study was to provide readers with a broad overview of the reporting and methodological characteristics of published Chinese observational study MAs. Although many such MAs have been published, the quality of these MAs is troubling. Thus, the reporting guidelines and methodological tools should be used to improve the quality of future MAs.

Author affiliations
¹School of Basic Medical Sciences, Lanzhou University, Lanzhou, China
²The First Hospital of Lanzhou University, Lanzhou, China
³The First School of Clinical Medicine of Lanzhou University, Lanzhou, China
⁴Evidence-Based Medicine Center, Institute of Traditional Chinese and Western Medicine, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

REFERENCES

1. Booth A, Clarke M, Ghersi D, et al. An international registry of systematic-review protocols. *Lancet* 2011;377:108–9.
2. Samaan Z, Mbuagbaw L, Kosa D, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *J of Multidiscip Healthc* 2013;6:169–88.
3. Mundi R, Chaudhry H, Singh I, et al. Checklists to improve the quality of the orthopaedic literature. *Indian J Orthop* 2008;42:150–64.
4. Sporbeck B, Jacobs A, Hartmann V, et al. Methodological standards in medical reporting. *J Dtsch Dermatol Ges* 2013;11:107–20.
5. Langan S, Schmitt J, Coenraads PJ, et al. The reporting of observational research studies in dermatology journals: a literature-based study. *Arch Dermatol* 2010;146:534–41.
6. Zhan SY. How to report systematic reviews and meta-analysis: introduction of QUOROM and MOOSE. *Chin J Evid Based Pediatr* 2010;5:60–3.
7. Xiong L, Du YH. Assessing tool of methodology quality for systematic reviews and meta-analysis: introduction of AMSTAR. *Chin J Evid Based Med* 2010;9:44.
8. Junhua, Z Hongcai, S Xiumei, G, et al. Methodology and reporting quality of systematic review/meta-analysis of traditional Chinese medicine. *J Altern Complement Med* 2007;13:797–805.
9. Ma B, Ke FY, Chen ZM, et al. Does the reporting of randomized clinical trials published in Chinese pediatrics journals improve after the CONSORT Statement is adopted? *Contemp Clin Trials* 2012;33:889–94.
10. Su N, LǔJ, Li C, et al. Assessment of reliability and validity of assessment of multiple systematic reviews in Chinese systematic reviews on stomatology. *Hua Xi Kou Qiang Yi Xue Za Zhi* 2013;31:49–52.
11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
12. Delaney A, Bagshaw SM, Ferland A, et al. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care* 2005;9:R575–82.
13. Ma B, Guo J, Qi G, et al. Epidemiology, quality and reporting characteristics of systematic reviews of traditional Chinese medicine interventions published in Chinese journals. *PLoS ONE* 2011;6: e20185.

14. von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
15. da Costa BR, Cevallos M, Altman DG, *et al*. Uses and misuses of the STROBE statement: bibliographic study. *BMJ Open* 2011;1: e000048.
16. Wells G, Shea B, O'Connell D, *et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis*. Ottawa: University of Ottawa, 2008. http://www.ohri. ca/programs/clinical-epidemiology/oxford-web.ppt
17. Critical Skills Appraisal Program (CASP) [EB/OL]. 2012. http://www. casp-uk.net/wp-content/uploads/2011/11/CASP-Cohort-Appraisal-Checklist-14oct10.pdf
18. Barnes DE, Bero LA. Why review article on health effects of passive smoking reach different conclusions. *JAMA* 1998;279:1566–70.
19. Groenwold RHH, Van Deursen AMM, Hoes AW, *et al*. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Ann Epidemiol* 2008;18:746–51.
20. Thompson SG. Systematic review: why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994;309:1351–5.
21. Papathanasiou AA, Zintzaras E. Assessing the quality of reporting of observational studies in cancer. *Ann Epidemiol* 2010;20:67–73.
22. Panic N, Leoncini E, de Belvis G, *et al*. Evaluation of the endorsement of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS ONE* 2013;8: e83138.