Journal of Clinical Epidemiology 107 (2019) 1−11

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

The methodological quality of dose-response meta-analyses needed substantial improvement: a cross-sectional survey and proposed recommendations

Chang Xu[a], Yu Liu[b], Peng-Li Jia[c], Ling Li[a], Tong-Zu Liu[d], Liang-Liang Cheng[a], Ke Deng[a], A.S.M. Borhan[e], Lehana Thabane[f], Xin Sun[a,*]

[a]*Chinese Evidence-Based Medicine Center and CREAT Group, West China Hospital, Sichuan University and Collaborative Innovation Center, Chengdu, China*

[b]*Gansu Provincial Maternity and Child-Care Hospital, Gansu, China*

[c]*School of Management, Shanxi Medical University, Taiyuan, China*

[d]*Department of Urology, Zhongnan Hospital, Wuhan University, Wuhan, China*

[e]*Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1; Biostatistics Unit, Father Sean O'Sullivan Research Centre, St Joseph's Healthcare, 3rd Floor, Martha Wing, Room H-325, 50 Charlton Avenue East, Hamilton, ON L8N 4A6, Canada*

[f]*Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1; Biostatistics Unit, Father Sean O'Sullivan Research Centre, St Joseph's Healthcare, 3rd Floor, Martha Wing, Room H-325, 50 Charlton Avenue East, Hamilton, ON L8N 4A6, Canada*

Accepted 6 November 2018; Published online 13 November 2018

Abstract

Objectives: To investigate methodological quality of published dose-response meta-analysis (DRMA) and explore study characteristics associated with the quality.

Study Design and Setting: We searched three databases for published DRMAs and used a modified AMSTAR (15 items) checklist to assess the methodological quality. We summarized the compliance of those DRMAs to the AMSTAR items and used multivariable regression analysis to explore the association between prespecified study characteristics with the overall methodological quality.

Results: We identified 529 DRMAs. Of the methodological quality items, six were well complied (80% or more) and six poorly complied (30% or fewer) by these DRMAs. The median score was nine points [first and third quartile: 7, 10] and only 64/529 had score over 10 points. Regression analysis suggested that studies with more authors (β = 0.19; 95% confidence interval [CI]: 0.05, 0.33), published more recently (β = 0.29; 95% CI: 0.21, 0.36), with financial support (β = 0.41; 95% CI: 0.13, 0.70), conducted by authors from European (other regions vs. European, β = −0.68; 95% CI: −1.05, −0.31) were associated with better methodological quality.

Conclusion: The methodological quality of published DRMAs was suboptimal. Substantial efforts are warranted to improve the quality, including developing methodology guideline, involving more methodological trained authors, and so forth.

Keywords: Dose-response meta-analysis; Methodological quality; Cross-sectional survey; Compliance rate; Multivariable regression; AMSTAR

1. Introduction

In epidemiological research, investigating the dose-response relationship between certain exposure and outcomes (ie, dose-response study) has become a topic of increasing interest [1−3]. Dose-response meta-analyses (DRMAs) systematically synthesize dose-specific findings from multiple studies to yield more precise estimates of putative dose-response effects [4−10]. In nature, DRMA is a meta-regression analysis that treats exposure (eg, sleep duration) as an independent variable against the outcome (eg, risk of death), as a dependent variable within each study (two-stage approach), or the whole population (one-stage approach) [7,8].

The two commonly used statistical approaches for DRMA include the generalized least squares for trend

Conflict of interests: The authors have no conflict of interest to report.

Funding: X.S. was supported by Natural Science Foundation of China (No. 71573183) and C.X. was supported by Doctoral Scholarship of Sichuan University.

* Corresponding author. Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, 37 Guo Xue Xiang, Chengdu 610041, China. Tel.: +86 18 980606047; Fax: +86 28 85423040.

E-mail address: sunx79@hotmail.com (X. Sun).

https://doi.org/10.1016/j.jclinepi.2018.11.007

What is new?

Key findings

- The methodological quality of published dose-response meta-analysis was suboptimal, although it was improved over time.
- Studies with more authors or receiving financial support have better methodological quality and those by authors from Europe have superior methodological than studies by authors from Asia-Pacific regions.

What this adds to what was known?

- No studies have systematically examined the methodological of published dose-response meta-analysis. Our study has offered a comprehensive review of methodological quality of existing dose-response meta-analyses and identified study characteristics associated with the quality.
- We proposed a brief recommendation with five tips on the analytic issues of DRMA to help further review authors to better conduct DRMAs.

What is the implication and what should change now?

- Our study clearly suggests that substantial efforts are warranted to improve the methodological quality of such meta-analyses; and that an optimal dose-response meta-analysis may be conducted with more experienced authors and with adequate funding.

(GLST) method based on a two-stage approach [7] and the robust-error meta-regression (REMR) based on a one-stage approach [8]. For the two-stage approach, the within-study dose-response relationship was first fitted by linear or nonlinear regression function (stage 1), then the regression coefficients across studies were combined by a multivariate fixed- or random-effect model (stage 2) [7]. For the one-stage approach, each study was considered as a cluster and the dose-response relationship was estimated across the whole study populations using the mixed-effect clustered regression [8]. Unlike traditional meta-regression (TMR) method, DRMA requires within-study dose-specific effects, for which at least two categories (one-stage: two categories, eg, sleep for 4−6 hours vs. 6−8 hours; two-stage: three categories, sleep for 4−6 hours/8−10 hours vs. 6−8 hours) of exposure dose is needed from each study. This allows establishing a more precise estimation on the dose-response relationship between interested exposure and outcome.

In the past several years, an increasing number of DRMAs have been published [11]; some already had an impact on clinical practice [12−14]. Because DRMAs are more sophisticated than traditional meta-analyses, appropriate use of statistical and nonstatistical methods is highly important. For example, one important statistical issue in DRMA is how to test the nonlinearity (or linearity) of the pooled dose-response relationship, which is generally difficult to understand when fitting a restricted cubic spline function, and is subject to misuse. Another important issue is the selection of trend approximating model for a proper exposure. Both of which could lead to improper conclusion and mislead clinical practice.

Previous studies have investigated the methodological quality of various systematic reviews and meta-analyses, including traditional meta-analysis, diagnostic meta-analysis, single-case meta-analysis, and network meta-analysis [15−19]. The findings suggested unsatisfactory validity of the design and conduct, which may be correlated with some characteristics such as journal reputation, financial support, and author region [15−19]. The credibility of the results and the grade of the evidence of the findings would be largely influenced by the poor design and conduct. Dose-response meta-analyses may also face the same issue. However, no efforts have been made to examine methodological quality of published DRMAs. Understanding how well those published DRMAs were conducted and investigating factors associated with quality would offer important insights for the study design, conduct, and interpretation of future studies, particularly because DRMAs carry important specific methodological features. In addition, the availability of such information is also critical for the development of methodological guidance for DRMAs [20].

Therefore, we conducted a comprehensive literature review to examine the methodological quality of published DRMAs and explore factors associated with the quality. We also offered recommendations for the analysis of a DRMA.

2. Methods

2.1. Eligibility criteria

We included published DRMAs based on aggregate data. We defined DRMA as a type of meta-analysis that combines dose-response estimates between an independent variable (eg, sleep duration) and an outcome variable (eg, all-cause mortality) from studies addressing a similar question [4−10]. We included those DRMAs of binary outcome only due to the fact that very few DRMAs of continuous outcomes were available [7]. We excluded conference abstracts, brief reports, and letters, as these reports contain very limited information regarding study reporting.

2.2. Literature search

We searched Medline, Embase, and Wiley online Library for published DRMAs from 1st January, 2011 to

31st July, 2017 without language restriction. The first literature search was conducted on 1st January, 2016, and the updated search was conducted on 1st August, 2017. We restricted our search to this period because the statistical methodologies were not well established until 2011. The search terms included "dose-response meta-analysis," "non-linear meta-regression," "meta-analysis of prospective studies," "meta-analysis of cohort studies," and "meta-analysis of observational studies." The full search strategies are provided in the Supplementary File.

2.3. Study screening

Two experienced investigators (C.X. and Y.L.) screened the title and abstract for potential eligibility. Then, they independently read the full texts of potentially eligible articles for final selection. Both the authors have more than 3 years' experience on DRMA, and one author (X.C) primarily developed the REMR model [8] for DRMA.

2.4. Data collection

Two methods-trained investigators (C.X. and P-L.J.) collected the following information regarding study characteristics: name of first author, region of first author, number of authors, year of publication, type of studies included (eg, cohort study or case-control study), number of studies included, number of database searched, regression model used for DRMA, funding information (funded, not funded, and not reported), and utilization of GRADE approach (Grades of Recommendation, Assessment, Development, and Evaluation) [21]. The authors checked the collected information, and any disagreements were solved by discussion.

2.5. Assessment of methodological quality

We used the AMSTAR 1.0 tool [22] to assess the methodological quality of included DRMAs because, during the study, the AMSTAR 2.0 was not yet released [23]. We slightly modified the AMSTAR checklist by disaggregating four items to facilitate the assessment, as they typically contain two methodological issues. These items included "Was there duplicate study selection and data extraction?," "Was a comprehensive literature search performed?," "Was a list of studies (included and excluded) provided?," and "Was the scientific quality of the included studies assessed and documented?" For instance, we disaggregated the item "was there duplicate study selection and data extraction?" into two—"was there duplicate study selection?" and "was there duplicate data extraction?" Thus, the modification resulted in 15 items (Table S1).

Because of the differences in statistical modeling between DRMAs and traditional meta-analyses, we developed specific response rules for the item "were the methods used to combine the findings of studies appropriate?" We responded NO if a DRMA (1) fitting dichotomous variables (eg, heart rate, parity) in a nonlinear model. This is inappropriate for the disjoint nature and nonconvex property of such type of variables [24], although this did not influence of the parameters estimation; (2) combined descriptive studies (eg, cross-sectional or ecological study) with analytical studies (eg, cohort study or case-control studies), which would lead to inverse causality; (3) incorrectly presented results. For instance, for a two-stage GLST DRMA model, no intercept was set for the regression that the dose-response curve and confidence interval (CI) was forced to start from (or pass) the origin, that is, the reference (eg, zero) [7]. If the curve and CI did not converge on the origin, the result was judged inappropriately presented, which generally indicate a lack of centering procedure or improper plotting code that may lead to wrong conclusions (authors) and informed decisions (readers); or (4) used incorrect method to test the nonlinearity. For example, some authors claimed that they used four knots for the spline function, but they only set the coefficient of the second spline to zero to test the nonlinearity [25]. This has no impact on the parameter estimation; however, it impacted the interpretation of the results (eg, misinterpreted nonlinear results to linear results).

For the modified AMSTAR checklist, each item was related to three options: yes, no, can't answer. When assessing the methodological quality for each DRMA, one point was given if the DRMA met the requirement (answer "yes") for that specific item or zero if not (answer "no"). For the situation that the information was "can't answer," we did not give the credit (ie, treat it as "no" in the analysis) [26]. The total score was 15 points, and study with a higher score means better methodological quality.

In our study, we assessed the methodological quality by one investigators (X.C), then checked by another (L.Y). If any disagreement occurred, they conducted discussion to achieve consensus. Before the formal assessment, we conducted a pilot testing by using 39 DRMAs, in which the two investigators independently assessed methodological quality. They achieved the Kappa statistic of 0.70. To further improve the agreement, the two investigators conducted additional training by discussing the reasons for disagreements. During the formal assessment, they were also required that each investigator should spend at least 30 minutes for a published DRMA, and they were allowed to assess at most 20 DRMAs each day.

2.6. Sample size estimation

We estimated the minimum sample size (N) needed for the proportion of studies adhere to the checklist items by the formula (1) below [27]. This was conducted to ensure that a sufficient power would be achieved when calculating the adherence rate. We set three plausible values of the proportion (p) (0.1, 0.3, 0.5) to estimate the required sample size using a 95% CI of the proportion of studies that adhere

to the full checklist, and a margin error (E) of 0.1 and 0.05. These assumptions generated the required sample sizes varying from 35 to 384, with proportion of 0.5 generating the largest sample size of 384 (Table S2).

$$N = \frac{1.96^2 p(1-p)}{E^2} \quad (1)$$

2.7. Statistical analysis

We used descriptive statistics to present study characteristics. The radar plot and bar plot were used to show the extent to which those DRMAs met the requirement of each checklist item. We judged that an item was well complied if 80% or more of the included DRMAs met the item or poor complied if met by less than 30% [28]. Otherwise, the item was judged moderately complied by the included studies.

To investigate the association of study characteristics with the methodological quality, we prespecified four study characteristics that were likely related to the quality of DRMA in similar studies [15−20,28−30]. They were a number of authors (≤4, 5−6, 7−8, and >8, according to the interquartiles), publication year (continuous), region of first author (European vs. Asia-Pacific or America), and funding (yes vs. no or not reported).

The total methodology score was fitted in a linear regression model against these variables. Given the potential heteroscedasticity (by White's test), we used weighted least square regression model to fit the data [31]. A robust variance estimates estimator, by treating each journal as a cluster (prespecified), was added to address clustering of articles published in the same journal [8,32]. The generalized estimating equation with robust variance, a method for estimating the average response over the population instead of individual, was used as sensitivity analysis, considering the potential correlations between total scores (same journal) [32]. This would generate a conservative variance making the statistical inference more robust. We checked the potential multicollinearity of independent variables by correlations and variance inflation factor.

All data were extracted by Excel 2013 (Microsoft, Washington) and analyzed by STATA 14.0 (STATA, College Station, TX, Serial number: 10,699,393), with alpha = 0.05 as the criterion for statistical significance.

3. Results

We initially obtained 7,061 records from the literature search. After removing duplicates and the title/abstract screening, 1,306 publications were eligible for the full-text screening process. Finally, 529 DRMAs were included (Fig. 1).

The 529 DRMAs were published in 174 difference academic journals. Among these DRMAs, 350 (66.2%, 95% CI: 62.0%, 70.1%) were conducted by authors from Asia region, 129 (24.4%, 95% CI: 20.9%, 28.2%) by authors from Europe, 47 (8.9%, 95% CI: 6.7%, 11.6%) by authors from North America, and 3 (0.6%, 95% CI: 0.2%, 1.8%) by authors from Australia. The number of DRMAs was increasingly published in the last 4 years; 394 (75.0%, 95% CI: 70.6%, 78.0%) DRMAs were published after 2014. Cohort study was the most common type of study design included in DRMAs (n = 318, 60.1%, 95% CI: 55.9%, 60.2%), and restricted cubic spline was the most frequently used trend approximating model (n = 295, 55.8%, 95% CI: 51.5%, 56.0%).

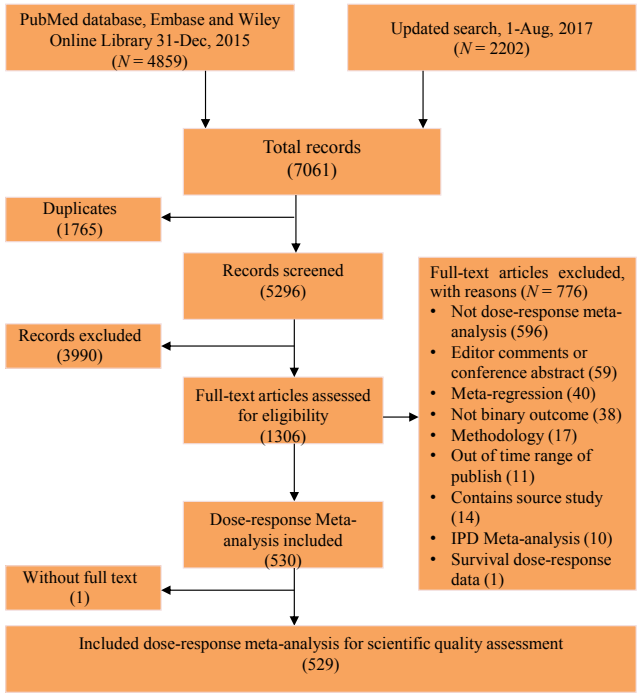

Fig. 1. The flow chart of literature screen. IPD, individual participant data.

The median number of authors was 6 [first quartile, third quartile: 4, 8], 32.3% (95% CI: 28.7%, 36.6%) of which have no more than four authors (n = 172). The median number of databases searched was 2 (first quartile, third quartile: 2, 3); 61 studies (11.5%, 95% CI: 9.1%, 14.5%) searched less than two databases. Only 6 (1.1%, 95% CI: 0.4%, 2.0%) of the DRMAs used the GRADE approach to rate the evidence. Most of the DRMAs received financial support (n = 337, 63.7%, 95% CI: 59.5%, 67.7%). For those receiving financial support, 336 were supported by government and only one by company. Table 1 presents the study details of the included DRMAs.

3.1. Adherence of the DRMAs to the modified AMSTAR checklist

Figure 2A and 2B present the level of adherence of the included DRMAs to each of the modified AMSTAR items. Of those 15 items, six were highly adhered by the included DRMAs (met by 80% or more of the DRMAs). These

Table 1. Basic characteristics of published DRMAs in past 7 years

Category by items	All publications (N = 529)
No. of authors (median [first and third quartiles])	6 (4 to 8)
≤4	171 (32.33%)
5 ~ 8	278 (52.55%)
>8	80 (15.12%)
Year of publication	
2011	35 (6.62%)
2012	44 (8.32%)
2013	56 (10.59%)
2014	117 (22.12%)
2015	120 (22.68%)
2016	85 (16.07%)
2017 (up to July 31)	72 (13.61%)
Database searched (median [first and third quartiles])	2 (2 to 3)
≤1	61 (11.53%)
2 ~ 3	385 (72.78%)
>3	83 (15.69%)
Journal distribution (n = 174 for journal numbers)	
Specialist journal (disease-specific)	365 (69.00%)
General journal (all diseases)	119 (22.50%)
Epidemiology or public health	45 (8.51%)
Design of source study	
Cohort	318 (60.11%)
Case-control	7 (1.32%)
Cross-section	3 (0.57%)
Mixed	199 (37.62%)
CCT and RCT	2 (0.38%)
No. of included studies (median [first and third quartiles])	14 (10 to 21)
≤10	151 (28.54%)
11 ~ 21	247 (46.69%)
>21	130 (24.57%)
Missing	1 (0.19%)
Region	
Asian	350 (66.16%)
European	129 (24.39%)
America	47 (8.88%)
Australia	3 (0.57%)
Model used in trend approximation[a]	
RCS regression	295 (55.77%)
FP regression	61 (11.53%)
Other nonlinear regression	21 (3.97%)
Linear	152 (28.73%)
Use of GRADE	
Yes	6 (1.13%)
No	523 (98.87%)
Funding	
Yes	337 (63.71%)
No	54 (10.21%)
Not reported	138 (26.09%)

Abbreviations: DRMA, dose-response meta-analysis; CCT, clinical controlled trial; RCT, randomized controlled trial.

[a] RCS, restricted cubic spline; FP, fractional polynomial; other nonlinear regression including natural cubic spline, quadratic polynomial, and so forth.

included search of at least two databases (89.0%, 95% CI: 86.1%, 91.4%), presentation of a list of included studies (96.8%, 95% CI: 94.9%, 98.0%), presentation of characteristics of included studies (98.3%, 95% CI: 96.8%, 99.1%), use of appropriate synthesis methods (86.0%, 95% CI: 82.8%, 88.7%), assessment of publication bias (95.7%, 95% CI: 93.6%, 97.1%), and statement of conflict of interest (92.6%, 95% CI: 90.1%, 94.6%).

The other six items were poorly adhered (met by less than 30% of the DRMAs), including provision of an a priori design (8.5%, 95% CI: 6.4%, 11.2%), documentation of search strategy (25.3%, 95% CI: 21.8%, 29.3%), use of publication status (gray literature) as inclusion criterion (24.2%, 95% CI: 20.7%, 28.1%), list of excluded studies (18.3%, 95% CI: 15.3%, 22.0%), documentation of scientific quality (28.2%, 95% CI: 24.5%, 32.3%), and appropriate use of scientific quality for conclusions (9.3%, 95% CI: 7.1%, 12.0%).

The remaining three items, which were moderately adhered, were use of duplicate study selection (46.9%, 95% CI: 42.7%, 51.1%), involvement of duplicate data extraction (77.7%, 95% CI: 74.0%, 81.0%), and assessment of scientific quality (57.1%, 95% CI: 52.8%, 61.2%).

3.2. Analytic issues of the synthesis methods

Figure 3 presented the detailed analytic issues of inappropriate use of synthesis method. There were 74 DRMAs (14.0%, 95% CI: 11.3%, 17.2%) that used at least one of the four inappropriate statistical models we defined earlier. Of which, 12 used dichotomous variables in a nonlinear model (16.2%, 95% CI: 8.7%, 26.6%), 19 combined descriptive studies with analytical studies (25.7%, 95% CI: 16.2%, 37.2%), 21 incorrectly presented results (28.4%, 95% CI: 18.5%, 40.1%), 13 used incorrect method to test the nonlinearity (17.6%, 95% CI: 9.7%, 28.2%), and nine mixed (all mixed with two of the errors) above errors (12.2%, 95% CI: 5.7%, 21.8%).

3.3. Study characteristics associated with methodological quality

The global methodological quality score ranged from 4 to 15, with the median of 9 (first and third quartiles: 7, 10). Only 64 (12.1%) of included studies had score higher

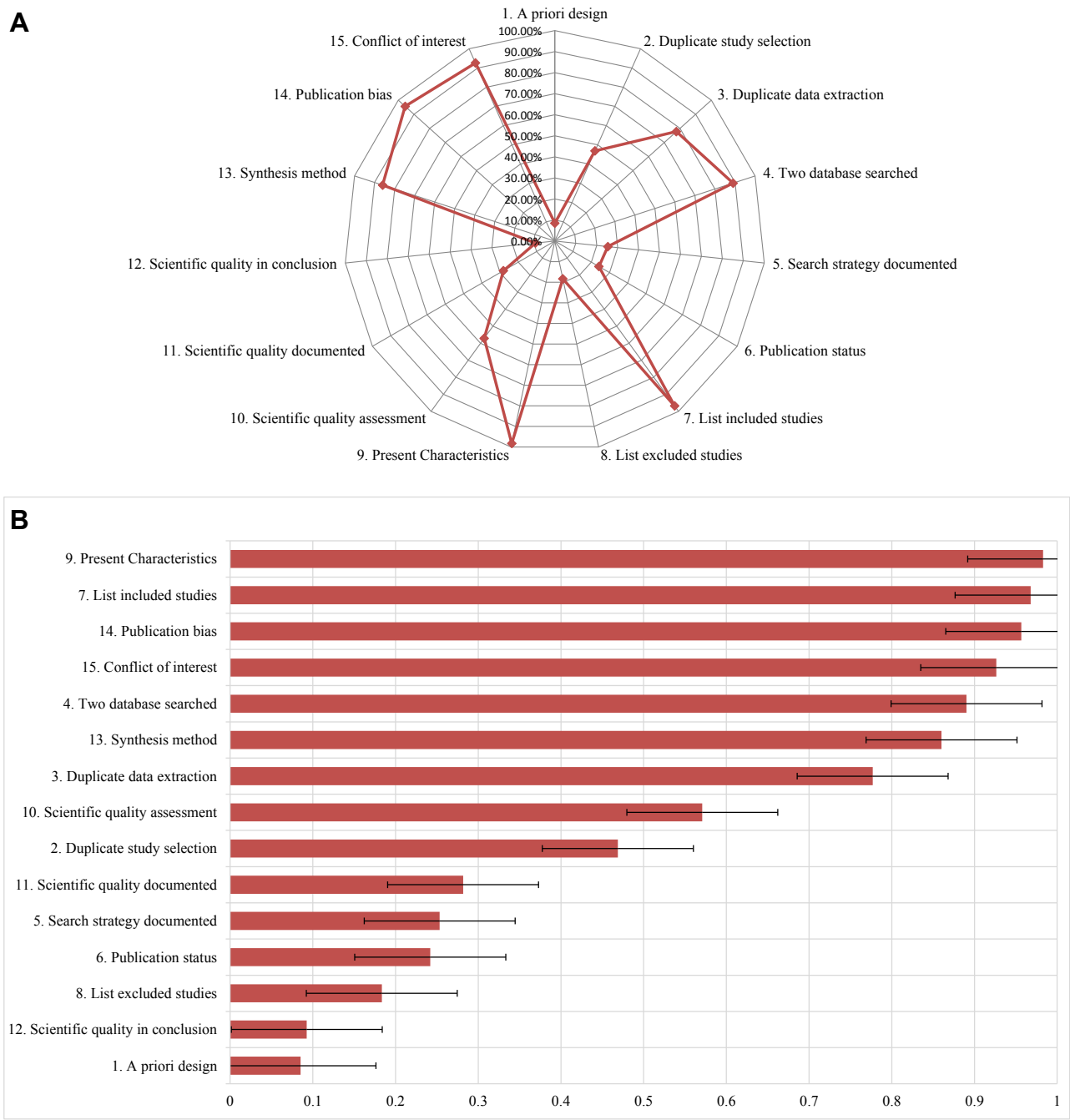

Fig. 2. The adherence rate of each methodological tip; (A) radar plot of the adherence rate of each item and (B) the ranking of the adherence rate.

than 10 points. Figure 4 presents the distribution of methodological quality score.

Our multivariable regression analysis suggested that, after adjusted for clustering on journal, a larger number of authors (5 to 6 vs. 4 or less [estimated β = 0.71; 95% CI: 0.39, 1.03; $P < 0.0001$]), studies published more recently (estimated β = 0.29; 95% CI: 0.21, 0.36; P for trend < 0.0001), studies with financial support (estimated β = 0.41; 95% CI: 0.13, 0.70; $P = 0.005$) were associated with better methodological quality, whereas studies from authors other than European region (estimated β = −0.68; 95% CI: −1.05, −0.31; $P < 0.0001$) were associated with poorer methodological quality (Table 2). The multivariable regression can explain 17.5% of the variance [$R^2 = 0.175$; F (13,173) = 7.99].

3.4. Sensitivity analysis

We used generalized estimating equation with clustered variance as a sensitivity analysis of the aforementioned multivariate regression analysis. This was achieved by setting the variable of "journal" as clustered variable. No

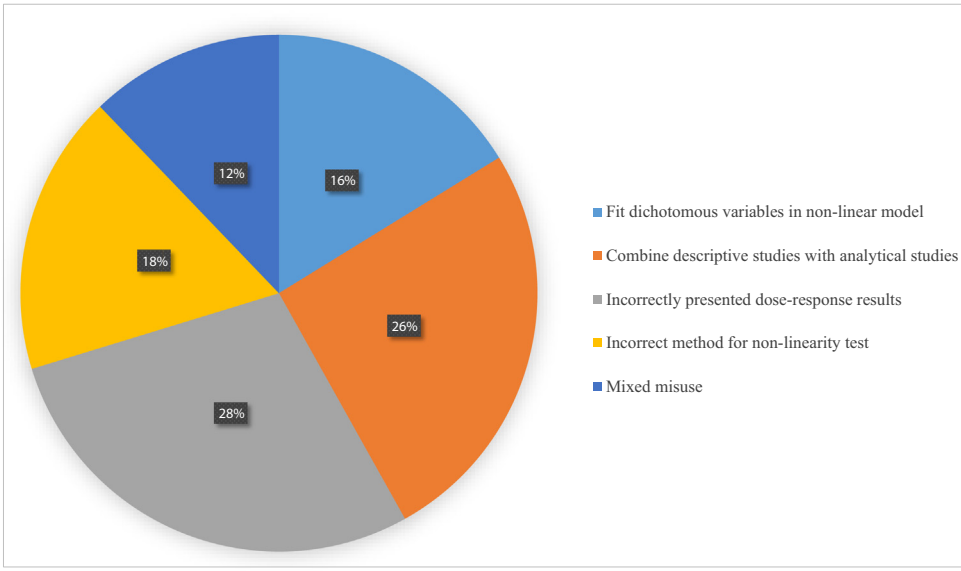

Fig. 3. The pie plot of analytic issues of synthesis methods of 74 DRMAs. DRMA, dose-response meta-analysis.

significant changes occurred, suggesting robustness of the findings (Table 2).

4. Discussion

In this comprehensive literature review, we found that the methodological quality of published DRMAs was suboptimal. The major limitations included the failures to provide an a priori design, document search strategy, consider gray literature as an inclusion criterion, list excluded studies, document scientific quality, and appropriately factor scientific quality into conclusions. In the regression analysis, we found that involvement of more authors and provision of financial support were associated with higher methodological quality of DRMA. We also observed significant improvement of the methodological quality over the recent years. This positive trend was similar to a recent research about meta-analyses of single-case experimental studies as well as two other surveys of meta-analysis on specialist areas (pediatric dentistry, vaccinology) [17,33,34].

In our study, DRMAs conducted by Asia-Pacific authors had lower methodological quality than those by European authors. The finding was consistent with some other meta-epidemiological studies [22,35,36]. In the study by Qin et al. [36], meta-analyses in English showed better methodological quality than Chinese. This suggested that more efforts may be warranted for Asian authors to use rigorous methods of DRMAs. Fortunately, research initiatives have been made to change this

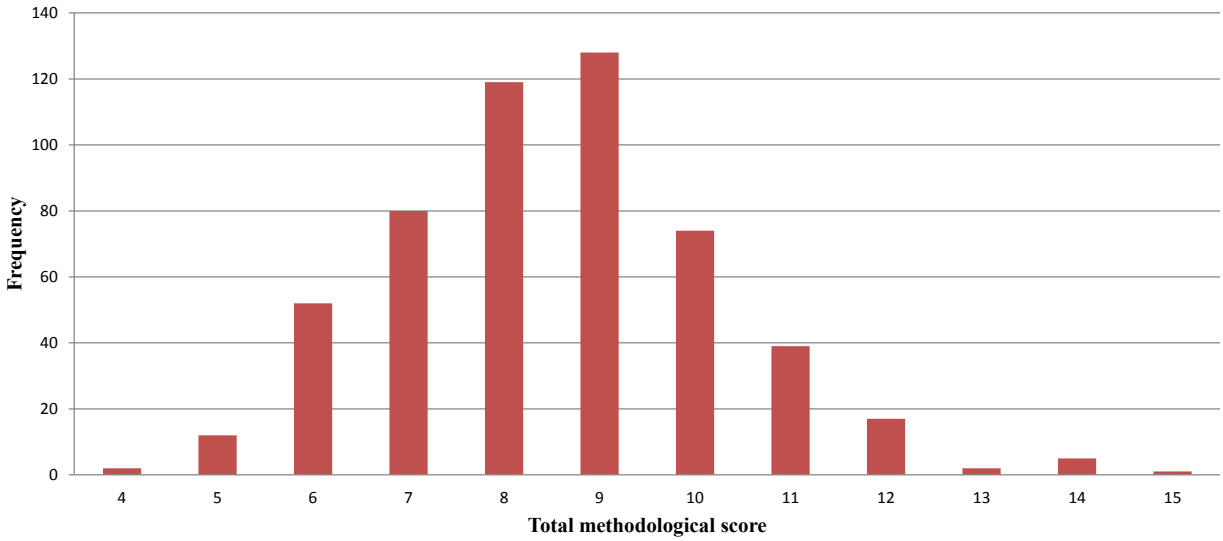

Fig. 4. The distribution of total methodology quality.

situation. For instance, a series of methodological articles have been available to the Chinese audiences, which specifically discussed the methods of DRMAs [37,38]. The reporting guidance of DRMAs (Chinese version) was also published in 2016 [11].

Methodological quality has close relevance to internal validity of meta-analysis. Poorly designed and conducted meta-analyses are susceptible to systematic bias, and the results are less useful for evidence-informed decision [39]. In this survey, the major methodological limitations of the included DRMAs lied in literature search and screening, and risk of bias assessment. Low reproducibility in literature search and use of less rigorous study screening processes may introduce selective bias. For the risk of bias assessment, the most common problem was that the authors simply presented a summary score of risk of bias, instead of documenting details. Such aggregate information is insufficient for clinical guideline developers to achieve informed decision. To improve the methodological quality of DRMAs, substantial efforts should be implemented urgently. Review authors are highly suggested to cooperate with methodologists to ensure the appropriate conduct of statistical and nonstatistical methods. Developing and publishing a protocol in advance may also be helpful to improve the design and conduct [40]. The GRADE approach is highly recommended to be used in DRMAs to rate the quality of evidence to achieve an informed decision for evidence users [21]. In addition, it is reasonable to develop a methodological guideline of DRMA to help review authors to form a clear thinking pathway.

Among the AMSTAR items, several were well adhered by most of the DRMAs. Nevertheless, the issue regarding the use of appropriate synthesis methods is worth discussing. An appreciable proportion of DRMAs (nearly 14%) failed to use appropriate methods to synthesize effect estimates from individual studies. It is particularly concerning that some studies combined effect estimates from analytical studies (eg, cohort studies) with descriptive studies (eg, cross-sectional studies) or used an inappropriate model we documented earlier [4−10]. These types of inappropriate use of the methods would largely influence the validity of the results and the conclusions of the findings that make no sense for the health care decision. It should be noted that this proportion is largely underestimated because the description of the synthesis methods of dose-response relationship is insufficient, most of which did not provide a clear description on how they fit the dose-response relationship, synthesize the dose-response relationship, and test the potential trend (eg, linear or nonlinear). We cannot make a thorough judgment of such studies of whether the

Table 2. Multivariable regression analysis of potential factors for methodological quality

	Estimated β (95% CI)		Sensitivity analysis	
Influence factors	**Multivariable**	***P*-value**	**GEE regression**	***P*-value**
No. of authors				
≤4	Reference		Reference	
5 ~ 6	0.71 (0.39, 1.03)	<0.0001	0.73 (0.41, 1.05)	<0.0001
7~ 8	0.72 (0.37, 1.08)	<0.0001	0.83 (0.45, 1.20)	<0.0001
>8	0.45 (0.02, 0.87)	0.051	0.57 (0.13, 1.00)	0.011
Each author increase	0.19 (0.05, 0.33)	0.010	0.24 (1.57, 3.06)	0.001
Year of publication				
2011	Reference		Reference	
2012	0.91 (0.24, 1.57)	0.007	0.89 (0.29, 1.49)	0.004
2013	1.14 (0.35, 1.93)	0.003	1.20 (0.48, 1.93)	0.001
2014	1.14 (0.49, 1.78)	0.001	1.15 (0.54, 1.76)	<0.0001
2015	1.43 (0.75, 2.11)	<0.0001	1.41 (0.76, 2.06)	<0.0001
2016	1.74 (1.07, 2.41)	<0.0001	1.75 (1.15, 2.36)	<0.0001
2017 (up to July 31)	2.26 (1.53, 3.01)	<0.0001	2.31 (1.57, 3.04)	<0.0001
Each year increase	0.29 (0.21, 0.36)	<0.0001	0.30 (0.22, 0.39)	<0.0001
Region				
European	Reference		Reference	
Asia-Pacific	−0.68 (−1.05, −0.31)	<0.0001	−0.77 (−1.13, −0.41)	<0.0001
America	−0.34 (−0.97, 0.28)	0.265	−0.41 (−1.07, 0.25)	0.217
Funding				
No or not reported	Reference		Reference	
Yes	0.41 (0.13, 0.70)	0.005	0.34 (0.05, 0.63)	0.021

Abbreviation: CI, confidence interval.

The multivariable regression was based on weighted least square linear regression; the sensitivity analysis was based on generalized estimating equation (GEE); both the methods with the variance estimation were based on robust standard error.

methods are appropriately used. Therefore, the current survey may overestimate the methodological quality of these DRMAs.

Referring to these analytic issues, there are five recommendations for conducting a DRMA.

4.1. To ensure what types of exposure variable will be fitted in the DRMA model

Continuous variable can be fitted in both linear and nonlinear regression functions, whereas for dichotomous variables, linear or piecewise linear models are recommended.

4.2. To ensure what types of study will be included and synthesized

Generally, it is improper to synthesize the results of descriptive studies (eg, cross-sectional) with analytic studies (eg, cohort, case-control) together to avoid the inverse causality [41]. However, it is not meant that pooling cohort studies (or case-control studies) can escape from inverse causality due to varying of confounders.

4.3. To ensure what kind of trend approximating function for fitting the potential nonlinear dose-response relationship

There were generally three types of functions for consideration: restricted cubic splines, natural quadratic function, and the fractional polynomials [4,6,10]. The most commonly used nonlinear function is the restricted cubic splines with 3 or 4 knots inserted in the data distribution. Inserting three knots would generate two splines with the first spline as dose itself (linear part), whereas the second is the cubic transformation (nonlinear part) [6,7]. When testing the nonlinearity, we can assume the coefficient of the second spline as zero to get a statistical inference. Similarity, inserting four knots would generate three splines with the second and third splines as nonlinear part, which were both assumed to be zero for the nonlinearity test [25].

4.4. To ensure what kind of approaches are used for synthesizing the dose-response relationship

TMR model, GLST, and REMR were all available approaches. TMR did not take into account for the within-study correlations of relative risks that the estimation is relatively crude [42]. GLST is generally a two-stage procedure while REMR is a one-stage procedure; empirical evidence suggested that these two methods can reach similar trend estimation [6−8].

4.5. To ensure whether the reference categories across studies are homogeneous

When there are heterogeneous references, centering the references is a necessary procedure to reduce the heterogeneity [8] because heterogeneity indicates some sort of inconsistency in the main effect, which is problematic for making clinical choices.

Our study has some strengths. In this survey, we included nearly all of the DRMAs published before August, 2017. Thus, our analyses have a high level of representativeness. Second, we used rigorous methods to conduct this study. For instance, we mandated a strict process of quality assessment, in which, we required that at least 30 minutes should be used for assessing methodological quality of each study. Third, we used appropriate statistical methods, such as robust weighted least square regression. To the best of our knowledge, this is a most comprehensive survey regarding DRMAs and the first investigation of methodological quality of these DRMAs. The findings would be useful for the development of methodology guideline of DRMAs [20].

There were also few limitations. First, we did not include DRMAs with continuous outcomes. Nevertheless, this type of DRMAs is very rare, and we decided not to include for the sake of interpretability of findings. We will continue our efforts to examine methodological quality of DRMAs of a continuous outcome when the data are sufficient. Second, we may not have identified all potential factors associated with methodological quality. For example, we were unable to acquire length of time spent on conducting each DRMA due to the lack of such information from each report. Third, in assessing the global methodological quality, we arbitrarily allocated one point to each of the 15 modified AMSTAR items, assuming that each item carried equal weight. This approach may not be optimal. However, the analyses using alternative approach suggested that the findings were robust. In addition, we used 80% and 30% as cut points for measurement of well complied, moderate complied, and poor complied; this was somewhat arbitrary that different cut point setup may lead to different judgment.

In conclusion, the methodological quality of published DRMAs was suboptimal, although it has been improved over time. More efforts are warranted for authors from Asia to take more rigorous methods in the conduct of the study. It would be highly desirable if more methodologically trained authors are involved in the study to ensure the methodological quality. In addition, funding bodies may consider offering support to this type of study to ensure sufficient resources are available to the investigators. An explicit guidance specific to the conduct of DRMAs may also be desirable to improve the quality of future DRMAs.

Acknowledgments

X.C. would like to express his deep appreciation for Prof. Suhail A.R Doi (Qatar University) for his guidance on him of synthesis methods for dose-response data.

Authors' contributions: X.S., T.Z.L., and C.X. conceived and designed the study; C.X. and Y.L. contributed to the quality assessment; P-L.J. and C.X. conducted the data collection; C.X. conducted the literature search, analyzed the data, plotted the figures and tables, and drafted the article; X.S. and L.T. provided statistical guidance; X.S., L.L., Y.L., T.L., L-L.C., P-L.J., K.D., and ASM.B. provided careful comments and revised the article. All authors approved the final version. The primary data can be obtained from the first author (C.X.).

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2018.11.007.

References

[1] Delcoigne B, Colzani E, Prochazka M, Gagliardi G, Hall P, Abrahamowicz M, et al. Breaking the matching in nested case-control data offered several advantages for risk estimation. J Clin Epidemiol 2017;82:79−86.

[2] Shadish WR, Rindskopf DM, Boyajian JG. Single-case experimental design yielded an effect estimate corresponding to a randomized controlled trial. J Clin Epidemiol 2016;76:82−8.

[3] Lauritzen HB, Larose TL, Øien T, Sandanger TM, Odland JØ, van de Bor M, et al. Prenatal exposure to persistent organic pollutants and child overweight/obesity at 5-year follow-up: a prospective cohort study. Environ Health 2018;17(1):9.

[4] Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose−response data, with an application to alcohol and mortality. Am J Epidemiol 2004;159:1077−86.

[5] Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. Epidemiology 1993;4:218−28.

[6] Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. Am J Epidemiol 2012;175:66−73.

[7] Crippa A, Orsini N. Dose-response meta-analysis of differences in means. BMC Med Res Methodol 2016;16:91.

[8] Xu C, Doi SA. The robust-error meta-regression method for dose-response meta-analysis. Int J Evid Based Healthc 2018;16(3):138−44.

[9] Dumouchel W. Meta-analysis for dose− response models. Stat Med 1995;14:679−85.

[10] Liu Q, Cook NR, Bergström A, Hsieh CC. A two-stage hierarchical regression model for meta-analysis of epidemiologic nonlinear dose−response data. Comput Stat Data Anal 2009;53:4157−67.

[11] Xu C, Doi SA, Zhang C, Sun X, Chen H, Zhou Q, et al. Proposed reporting guidelines for dose− response meta-analysis (Chinese Edition). Chin J Evid Based Med 2016;16:1221−6.

[12] Liu TZ, Xu C, Matteo R, Cai H, Zhang C, Shi MJ, et al. Sleep duration and risk of all-cause mortality: a flexible non-linear meta-regression of 40 prospective cohort studies. Sleep Med Rev 2017;32:28−36.

[13] Crippa A, Larsson SC, Discacciati A, Wolk A, Orsini N. Red and processed meat consumption and risk of bladder cancer: a dose-response meta-analysis of epidemiological studies. Eur J Nutr 2018;57:689−701.

[14] Rota M, Pasquali E, Bellocco R, Bagnardi V, Scotti L, Islami F, et al. Alcohol drinking and cutaneous melanoma risk: a systematic review and dose-risk meta-analysis. Br J Dermatol 2014;170:1021−8.

[15] Adie S, Ma D, Harris IA, Naylor JM, Craig JC. Quality of conduct and reporting of meta-analyses of surgical interventions. Ann Surg 2015;261(4):685−94.

[16] Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Cochrane diagnostic test accuracy working group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889−97.

[17] Jamshidi L, Heyvaert M, Declercq L, Fernández-Castilla B, Ferron JM4, Moeyaert M, et al. Methodological quality of meta-analyses of single-case experimental studies. Res Dev Disabil 2018;79:97−115.

[18] Bafeta A, Trinquart L, Seror R, Ravaud P. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. BMJ 2013;347:f3675.

[19] Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Systematic review adherence to methodological or reporting quality. Syst Rev 2017;6(1):131.

[20] Selva A, Sanabria AJ, Pequeño S, Zhang Y, Solà I, Pardo-Hernandez H, et al. Incorporating patients' views in guideline development: a systematic review of guidance documents. J Clin Epidemiol 2017;88:102−12.

[21] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383−94.

[22] 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;7:10.

[23] 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.

[24] Arora JS, Huang MW, Hsieh CC. Methods for optimization of nonlinear problems with discrete variables: a review. Struct Optim 1994;8:69.

[25] Durrleman S, Simon R. Flexible regression models with cubic splines. Stat Med 1989;8:551−61.

[26] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408−12.

[27] Abbade LP, Wang M, Sriganesh K, Mbuagbaw L, Thabane L. Framing of research question using the PICOT format in randomised controlled trials of venous ulcer disease: a protocol for a systematic survey of the literature. BMJ Open 2016;6(11):e013175.

[28] Jia PL, Tang L, Yu JJ, Liu J, Kang D, Sun X. The quality of reporting in randomized controlled trials of acupuncture for knee osteoarthritis: a cross-sectional survey. PLoS One 2018;13:e0195652.

[29] Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA 2001;285:437−43.

[30] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609−13.

[31] Eisenhauer JG. Regression through the origin. Teach Stat 2003;25:76−80.

[32] Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. Int Stat Rev 1991;59(1):25−35.

[33] Jayaraman J, Nagendrababu V, Pulikkotil SJ, Innes NP. Critical appraisal of methodological quality of systematic reviews and meta-analysis in paediatric dentistry journals. Int J Paediatr Dent 2018;28:548−60.

[34] De Vito C, Manzoli L, Marzuillo C, Anastasi D, Boccia A, Villari P. A systematic review evaluating the potential for bias and the methodological quality of meta-analyses in vaccinology. Vaccine 2007;25(52):8794−806.

[35] Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 2016; 94(3):485−514.

[36] Qin Y, Yao L, Shao F, Yang K, Tian L. Methodological quality assessment of meta-analyses of hyperthyroidism treatment. Horm Metab Res 2018;50(1):8−16.

[37] Xu C, Zhang YG, Han FF, Niu YM, Kuang XY, Zhang C. How to perform dose-response meta-analysis: a brief introduction of methodology. Chin J Evid Based Med 2015;15:1236−9.

[38] Xu C, Liu TZ, Kuang XY, Zhang YG, Weng H, Zhang C. How to estimate the missing data and transform the effect measure in dose-response meta-analysis. Chin J Evid Based Med 2015;15:984−7.

[39] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651−7.

[40] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;350:g7647.

[41] Shahar E, Shahar DJ. Causal diagrams and the cross-sectional study. Clin Epidemiol 2013;5:57−65.

[42] Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21:1559−73.